2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Distributed Convex Optimization with Identical Constraints

Shahin Nikookhoy, Jie Lu, and Choon Yik Tang

Abstract— **This paper presents a gossip-style, distributed asynchronous algorithm that solves constrained optimization problems over networks with time-varying topologies, where the objective function is a sum of uniformly strictly convex local objective functions belonging to nodes in the network, and the inequality and equality constraint functions are convex and identical to every node. Referred to as *Pairwise Equalizing* (PE), the algorithm operates by forcing the nodes' estimates of the unknown minimizer to asymptotically achieve consensus while satisfying a conservation condition derived from the Karush-Kuhn-Tucker condition. We show that as long as the gossiping pattern is sufficiently rich, PE achieves asymptotic convergence and solves the problem. The proposed algorithm represents an alternative to the existing subgradient algorithms and generalizes our earlier algorithm for problems without constraints.**

I. Introduction

In many envisioned applications of multi-agent systems, ad hoc networks, and sensor networks, nodes in the network have to accomplish tasks that require extensive processing of information, rapid decentralized decision making, and precise coordination of actions. Often, to achieve such goals, they have to distributively solve a convex optimization problem of the form

$$\min_{x\in\Omega}\sum_{i=1}^{N} f_i(x), \tag{1}$$

where f_i represents a convex local objective function observed by node i, Ω represents a convex constraint set known to every node, and N denotes the number of nodes. Since each node i knows only its own f_i, all of them must collaborate to solve problem (1) for the unknown minimizer x^*.

The current literature offers a collection of distributed algorithms for solving problem (1), which may be roughly classified into two groups. The first group contains the subgradient algorithms [1]–[17], which have a number of variations. For instance, a subset of the subgradient algorithms [1], [4]–[6], [11]–[14], [17] operate *incrementally*, relying on the passing of an estimate of x^* around the network to operate. Another subset of them [2], [3], [7]–[10], [15], [16] operate *non-incrementally*, with which each node maintains an estimate of x^* and updates it iteratively by exchanging information with neighbors. Regardless of the categories, the algorithms require stepsizes to operate. In addition, if the problem is constrained (i.e., the constraint set Ω is not the entire space), such algorithms also require projection during their operation. Moreover, in this case, different assumptions have been imposed concerning properties of the constraint set Ω. More specifically, [2], [4]–[6], [11], [13], [14], [16], [17] assume that Ω is nonempty, closed, and convex; [12] assumes further that it is compact; and [1] assumes that it is convex with a nonempty interior. Furthermore, in a recent work [10], the nodes are allowed to have non-identical constraint sets (i.e., each node i has a constraint set Ω_i).

The second group of distributed algorithms intended to solve problem (1) is made up of algorithms we developed, which are non-gradient-based [18], [19]. These algorithms were developed based on the ideas of satisfying a *conservation condition* and a *dissipation condition*. More specifically, with each of these algorithms, every node i in the network maintains an estimate x_i of x^* and updates it in such a manner that the sum of the gradients of the local objective functions, evaluated at these estimates, are *conserved* at zero, i.e., $\sum_{i=1}^N \nabla f_i(x_i)=0$—hence the name *conservation condition*. In addition, all the nodes update the estimates x_i's so that they gradually *dissipate* their differences, asymptotically achieving some consensus as the number of iterations approaches infinity—hence the name *dissipation condition*. For a problem without constraints, simultaneously satisfying these two conditions implies that all the x_i's must converge to the unknown minimizer x^*. It is based on these ideas that our algorithms in [18], [19] were developed. Unfortunately, these algorithms are limited only to problems *without* constraints.

In this paper, we show that the ideas of conservation and dissipation used in our earlier work can be extended to a class of problems *with* identical constraints, namely, problem (1) where each f_i is uniformly strictly convex, and Ω is closed, convex, and such that every point in Ω is a regular point. We present a gossip-style, distributed asynchronous algorithm, referred to as *Pairwise Equalizing* (PE), which solves problem (1) over networks with time-varying topologies, and which may be regarded as a generalization of one of our gossip algorithms for unconstrained problems reported in [18]. With PE, the nodes' estimates x_i's are forced to achieve consensus while satisfying a new conservation condition derived from the Karush-Kuhn-Tucker (KKT) condition. Using a Lyapunov-like function based on the Lagrangian of the problem, we show that as long as the gossiping pattern is sufficiently rich, PE is asymptotically convergent, driving all the x_i's to x^* and solving problem (1). Finally, we show that PE does not require stepsizes nor projection to operate, but does require the solving of a pairwise local optimization problem at each iteration, which may be computationally expensive.

The outline of this paper is as follows: Section II formulates the problem. Section III describes PE, while Section IV establishes its asymptotic convergence. Finally, Section V concludes the paper. Throughout the paper, let $\mathbb{P}=\{1,2,\ldots\}$, $\mathbb{P}_n=\{1,2,\ldots,n\}$, $\mathbb{N}=\{0,1,2,\ldots\}$, and

S. Nikookhoy, J. Lu, and C. Y. Tang are with the School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA (e-mail: {sh.nikookhoy, jie.lu-1, cytang}@ou.edu).

This work was supported by the National Science Foundation under grant CMMI-0900806.

∇ denote the gradient. In addition, for any vector $x \in \mathbb{R}^m$, we write $x = (x^{(1)}, x^{(2)}, \ldots, x^{(m)})$.

II. Problem Formulation

Consider a multi-hop network consisting of $N \geq 2$ nodes, connected by bidirectional links in a time-varying topology. The network is modeled as an undirected graph $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$, where $k \in \mathbb{N}$ denotes time, $\mathcal{V} = \{1, 2, \ldots, N\}$ represents the set of N nodes, and $\mathcal{E}(k) \subset \{\{i,j\} : i,j \in \mathcal{V}, i \neq j\}$ represents the nonempty set of links at time k. Any two nodes $i, j \in \mathcal{V}$ are one-hop neighbors and can communicate at time $k \in \mathbb{N}$ if and only if $\{i,j\} \in \mathcal{E}(k)$.

Suppose, at time $k = 0$, each node $i \in \mathcal{V}$ observes functions $f_i, g^{(1)}, ..., g^{(p)}, h^{(1)}, ..., h^{(m)} : \mathbb{R}^n \to \mathbb{R}$. Also suppose, upon observing the functions, all the N nodes wish to solve the following constrained, separable, optimization problem:

$$\min_{x \in \Omega} F(x), \tag{2}$$

where $F : \mathbb{R}^n \to \mathbb{R}$ is defined as $F(x) = \sum_{i \in \mathcal{V}} f_i(x)$ and

$$\Omega = \{x \in \mathbb{R}^n : g^{(\ell)}(x) \leq 0 \; \forall \ell \in \mathbb{P}_p, \; h^{(\ell)}(x) = 0 \; \forall \ell \in \mathbb{P}_m\}.$$

Moreover, suppose the following assumption holds. To state the assumption, consider the definition below:

Definition 1. A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *uniformly strictly convex* if there exists a continuous and strictly increasing function $\gamma : [0, \infty) \to [0, \infty)$ satisfying $\gamma(0) = 0$ and $\lim_{d \to \infty} \gamma(d) = \infty$ such that $f(y) - f(x) - \nabla f(x)^T (y - x) \geq \gamma(\|y - x\|) \; \forall x, y \in \mathbb{R}^n$.

Remark 1. Notice that uniform strict convexity is weaker than strong convexity (e.g., the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^4$ is uniformly strictly convex but not strongly convex) but stronger than strict convexity (e.g., $f(x) = e^{-x} + x$ is strictly convex but not uniformly strictly convex).

Assumption 1. The functions $f_i \; \forall i \in \mathcal{V}$, $g^{(1)}, ..., g^{(p)}$, $h^{(1)}, ..., h^{(m)}$ satisfy the following:

1) For each $i \in \mathcal{V}$, f_i is uniformly strictly convex and continuously differentiable.
2) For each $\ell \in \mathbb{P}_p$, $g^{(\ell)}$ is convex and continuously differentiable.
3) For each $\ell \in \mathbb{P}_m$, $h^{(\ell)}$ is affine.
4) The set Ω is nonempty.
5) For each $x \in \Omega$, x is a regular point [20], i.e., the vectors $\nabla g^{(\ell)}(x) \; \forall \ell \in J(x)$ and $\nabla h^{(\ell)}(x) \; \forall \ell \in \mathbb{P}_m$ are linearly independent, where $J(x) = \{\ell \in \mathbb{P}_p : g^{(\ell)}(x) = 0\}$ is the index set of the active inequality constraints.

Having introduced Assumption 1, we now show that problem (2) is well-posed:

Lemma 1. *For any uniformly strictly convex and continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and any nonempty closed convex set $D \subset \mathbb{R}^n$, there exists a unique $z^* \in D$ such that $f(z^*) \leq f(x) \; \forall x \in D$.*

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a uniformly strictly convex and continuously differentiable function and $D \subset \mathbb{R}^n$ be a nonempty closed convex set. Then, there exists a continuous and strictly increasing function $\tilde{\gamma} : [0, \infty) \to [0, \infty)$ satisfying $\tilde{\gamma}(0) = 0$ and $\lim_{d \to \infty} \tilde{\gamma}(d) = \infty$ such that $f(y) - f(x) - \nabla f(x)^T (y - x) \geq \tilde{\gamma}(\|y - x\|) \; \forall x, y \in \mathbb{R}^n$. Let $x_0 \in D$ and $D_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$. Note that $D_0 \neq \emptyset$. We first show that D_0 is convex and compact. Since f is convex, D_0 is convex and closed. To show that D_0 is bounded, pick any $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^n$ with $\|z\| = 1$. Because f is uniformly strictly convex, we have $\forall \eta \in \mathbb{P}$, $f(y + \eta z) - f(y + (\eta - 1)z) - \nabla f(y + (\eta - 1)z)^T z \geq \tilde{\gamma}(1)$ and $f(y + (\eta - 1)z) - f(y + \eta z) + \nabla f(y + \eta z)^T z \geq \tilde{\gamma}(1)$. Adding the two inequalities yields $\nabla f(y + \eta z)^T z - \nabla f(y + (\eta - 1)z)^T z \geq 2\tilde{\gamma}(1)$ $\forall \eta \in \mathbb{P}$. This, along with the fact that $\sum_{\eta=1}^{\zeta} \nabla f(y + \eta z)^T z - \nabla f(y + (\eta - 1)z)^T z = \nabla f(y + \zeta z)^T z - \nabla f(y)^T z$ $\forall \zeta \in \mathbb{P}$, implies that $\nabla f(y + \zeta z)^T z - \nabla f(y)^T z \geq 2\zeta \tilde{\gamma}(1)$ $\forall \zeta \in \mathbb{P}$. Hence, $\lim_{\zeta \to \infty} \nabla f(y + \zeta z)^T z = \infty$ and, thus, $\lim_{\|x\| \to \infty} \nabla f(x)^T z = \infty$. Since $\nabla f(x)^T z$ is the directional derivative of f along z at x, D_0 must be bounded. Therefore, D_0 is convex and compact. Because D is nonempty, convex, and closed, $D \cap D_0$ is nonempty, convex, and compact. Thus, due to the strict convexity and continuity of f, there exists a unique $z^* \in D \cap D_0$ such that $f(z^*) \leq f(x) \leq f(x_0)$ $\forall x \in D \cap D_0$. Since $f(x) > f(x_0) \; \forall x \in D - D_0$, z^* satisfies $f(z^*) \leq f(x) \; \forall x \in D$ and is unique. ☐

Proposition 1. *With Assumption 1, there exists a unique $x^* \in \Omega$, which minimizes F over Ω and solves problem* (2)*, i.e., $x^* = \arg\min_{x \in \Omega} F(x)$.*

Proof. By Assumption 1, F is uniformly strictly convex and continuously differentiable and Ω is nonempty, convex, and closed. It follows from Lemma 1 that there exists a unique $x^* \in \Omega$ that minimizes F over Ω, solving (2). ☐

Given the above, the goal of this paper is to construct a distributed asynchronous iterative algorithm with which every node is able to asymptotically determine the unknown minimizer x^*.

III. Pairwise Equalizing

In this section, we develop a gossip algorithm that solves problem (2).

To this end, note from Proposition 1 that with Assumption 1, there exist Lagrange multipliers $\mu^* \in \mathbb{R}^p$ and $\lambda^* \in \mathbb{R}^m$ of problem (2) such that x^*, μ^*, and λ^* uniquely satisfy the Karush-Kuhn-Tucker (KKT) conditions [21], i.e.,

$$\nabla_x \mathcal{L}(x^*, \mu^*, \lambda^*) = 0, \tag{3}$$

$$x^* \in \Omega, \tag{4}$$

$$\mu^{*(\ell)} \geq 0, \quad \forall \ell \in \mathbb{P}_p, \tag{5}$$

$$\mu^{*(\ell)} g^{(\ell)}(x^*) = 0, \quad \forall \ell \in \mathbb{P}_p, \tag{6}$$

where the Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ is defined as

$$\mathcal{L}(x, \mu, \lambda) = F(x) + N \sum_{\ell=1}^{p} \mu^{(\ell)} g^{(\ell)}(x) + N \sum_{\ell=1}^{m} \lambda^{(\ell)} h^{(\ell)}(x).$$

Note that the two N's in the Lagrangian $\mathcal{L}$ are introduced only for later convenience.

Suppose, at time $k = 0$, each node $i \in \mathcal{V}$ creates state variables $x_i \in \mathbb{R}^n$, $\mu_i = (\mu_i^{(1)}, \mu_i^{(2)}, \ldots, \mu_i^{(p)}) \in \mathbb{R}^p$, and $\lambda_i = (\lambda_i^{(1)}, \lambda_i^{(2)}, \ldots, \lambda_i^{(m)}) \in \mathbb{R}^m$, representing its estimate of x^*, μ^*, and λ^*, respectively. Also suppose, at each subsequent time $k \in \mathbb{P}$, an iteration, referred to as *iteration* k, occurs. Here we use $x_i(0)$ $\mu_i(0)$, and $\lambda_i(0)$ to denote the initial value of x_i, μ_i, and λ_i, and $x_i(k)$ $\mu_i(k)$, and $\lambda_i(k)$ their values upon completing iteration $k \in \mathbb{P}$. With the above, the aim of the algorithm may be stated as

$$\lim_{k\to\infty} x_i(k) = x^*, \quad \forall i \in \mathcal{V}. \tag{7}$$

To satisfy (7), consider a *conservation condition*, defined as

$$\sum_{i\in\mathcal{V}} \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) = 0, \quad \forall k \in \mathbb{N}, \tag{8}$$

$$x_i(k) \in \Omega, \quad \forall k \in \mathbb{N},\ \forall i \in \mathcal{V}, \tag{9}$$

$$\mu_i^{(\ell)}(k) \ge 0, \quad \forall \ell \in \mathbb{P}_p,\ \forall k \in \mathbb{N},\ \forall i \in \mathcal{V}, \tag{10}$$

$$\mu_i^{(\ell)}(k) g^{(\ell)}(x_i(k)) = 0, \quad \forall \ell \in \mathbb{P}_p,\ \forall k \in \mathbb{N},\ \forall i \in \mathcal{V}, \tag{11}$$

where $\mathcal{L}_i : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ is defined as

$$\mathcal{L}_i(x, \mu, \lambda) = f_i(x) + \sum_{\ell=1}^{p} \mu^{(\ell)} g^{(\ell)}(x) + \sum_{\ell=1}^{m} \lambda^{(\ell)} h^{(\ell)}(x).$$

Note that $\sum_{i\in\mathcal{V}} \mathcal{L}_i(x, \mu, \lambda) = \mathcal{L}(x, \mu, \lambda)$, which explains why the two N's are inserted into the Lagrangian $\mathcal{L}$. In addition, consider a *dissipation condition*, defined as

$$\lim_{k\to\infty} x_i(k) = \tilde{x}, \quad \forall i \in \mathcal{V},\ \text{for some } \tilde{x} \in \mathbb{R}^n. \tag{12}$$

The following theorem says that to achieve (7) in which the unknown minimizer x^* explicitly appears, it is sufficient to satisfy both the conservation condition (8)–(11) and the dissipation condition (12), which do not contain x^*:

Theorem 1. *Suppose Assumption 1 holds. Then, with* (8)–(11) *and* (12)*,* $\tilde{x} = x^*$*, i.e.,* (7) *holds.*

Proof. Due to space limitations, the proof is omitted. □

Suggested by Theorem 1, we now construct a gossip algorithm that satisfies the conservation condition (8)–(11) and *attempts* to satisfy the dissipation condition (12). To ensure the conservation condition, observe that (8) holds if and only if

$$\nabla_x \mathcal{L}_i(x_i(0), \mu_i(0), \lambda_i(0)) = 0, \quad \forall i \in \mathcal{V} \tag{13}$$

and

$$\sum_{i\in\mathcal{V}} \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) = \sum_{i\in\mathcal{V}} \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)), \quad \forall k \in \mathbb{P}. \tag{14}$$

The proposition below shows that (13), together with (9)–(11), uniquely determines the initial states $x_i(0)$, $\mu_i(0)$, and $\lambda_i(0)$ $\forall i \in \mathcal{V}$:

Proposition 2. *Suppose Assumption 1 holds. Then, for each* $i \in \mathcal{V}$*, there exists a unique* $x_i^* \in \Omega$ *that minimizes* f_i *over* Ω*, i.e.,* $x_i^* = \arg\min_{x\in\Omega} f_i(x)$*. Moreover, there exist* $\mu_i^* \in \mathbb{R}^p$ *and* $\lambda_i^* \in \mathbb{R}^m$ *such that* x_i^**,* μ_i^**, and* λ_i^* *uniquely satisfy the following KKT conditions:*

$$\nabla_x \mathcal{L}_i(x_i^*, \mu_i^*, \lambda_i^*) = 0, \tag{15}$$

$$x_i^* \in \Omega, \tag{16}$$

$$\mu_i^{*(\ell)} \ge 0, \quad \forall \ell \in \mathbb{P}_p, \tag{17}$$

$$\mu_i^{*(\ell)} g^{(\ell)}(x_i^*) = 0, \quad \forall \ell \in \mathbb{P}_p. \tag{18}$$

Proof. It follows from Assumption 1 and Lemma 1 that for each $i \in \mathcal{V}$, there exists a unique $x_i^* = \arg\min_{x\in\Omega} f_i(x)$. It follows from [21] that there exist μ_i^* and λ_i^* such that x_i^*, μ_i^*, and λ_i^* uniquely satisfy (15)–(18). □

Proposition 2 asserts that to satisfy (13) and (9)–(11) for $k = 0$, it suffices that each node $i \in \mathcal{V}$ finds the unique minimizer x_i^* and the unique Lagrange multipliers μ_i^* and λ_i^* of the constrained convex optimization problem $\min_{x\in\Omega} f_i(x)$, and then sets its initial states as follows:

$$x_i(0) = x_i^*, \tag{19}$$

$$\mu_i(0) = \mu_i^*, \tag{20}$$

$$\lambda_i(0) = \lambda_i^*. \tag{21}$$

Next, to satisfy (14) and (9)–(11) for $k \in \mathbb{P}$, suppose at each iteration $k \in \mathbb{P}$, a pair $u(k) = \{u_1(k), u_2(k)\} \in \mathcal{E}(k)$ of one-hop neighbors $u_1(k)$ and $u_2(k)$ gossip and update their state variables, while the remaining nodes stay idle, i.e.,

$$x_i(k) = x_i(k-1), \quad \forall k \in \mathbb{P},\ \forall i \in \mathcal{V} - u(k), \tag{22}$$

$$\mu_i(k) = \mu_i(k-1), \quad \forall k \in \mathbb{P},\ \forall i \in \mathcal{V} - u(k), \tag{23}$$

$$\lambda_i(k) = \lambda_i(k-1), \quad \forall k \in \mathbb{P},\ \forall i \in \mathcal{V} - u(k). \tag{24}$$

Thus, with (22)–(24), (14) becomes

$$\sum_{i\in u(k)} \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) = \sum_{i\in u(k)} \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)), \quad \forall k \in \mathbb{P}. \tag{25}$$

Hence, (14) can be satisfied simply through a gossip between nodes $u_1(k)$ and $u_2(k)$ to share their observed functions and state variables, which allows the two nodes to jointly update their state variables according to (25).

Observe that (25) does not uniquely determine $x_{u_1(k)}(k)$, $x_{u_2(k)}(k)$, $\mu_{u_1(k)}(k)$, $\mu_{u_2(k)}(k)$, $\lambda_{u_1(k)}(k)$, and $\lambda_{u_2(k)}(k)$. Therefore, we may use the available degree of freedom to account for the dissipation condition (12). Since (12) requires all the $x_i(k)$'s to achieve consensus and since it is desirable that all the $\mu_i(k)$'s and $\lambda_i(k)$'s remain bounded, imposing the following *equalizing condition* may make these happen:

$$x_{u_1(k)}(k) = x_{u_2(k)}(k), \quad \forall k \in \mathbb{P}, \tag{26}$$

$$\mu_{u_1(k)}(k) = \mu_{u_2(k)}(k), \quad \forall k \in \mathbb{P}, \tag{27}$$

$$\lambda_{u_1(k)}(k) = \lambda_{u_2(k)}(k), \quad \forall k \in \mathbb{P}. \tag{28}$$

The following proposition says that (25), (26)–(28), and (9)–(11) collectively have a unique solution, so that the evolutions of the $x_i(k)$'s, $\mu_i(k)$'s, and $\lambda_i(k)$'s are well-defined:

Proposition 3. *Suppose Assumption 1 holds. Then, with* (19)–(21)*,* (22)–(24)*,* (25)*,* (26)–(28)*, and* (9)–(11)*,* $x_i(k)$*,* $\mu_i(k)$*,* $\lambda_i(k)$ $\forall k \in \mathbb{N}$ $\forall i \in \mathcal{V}$ *are well-defined.*

Proof. By induction on $k \in \mathbb{N}$. Because of Assumption 1, (19)–(21), and Proposition 2, $x_i(0)$, $\mu_i(0)$, $\lambda_i(0)$ $\forall i \in \mathcal{V}$ are well defined. Next, let $k \in \mathbb{P}$ and suppose $x_i(k-1)$, $\mu_i(k-1)$, $\lambda_i(k-1)$ $\forall i \in \mathcal{V}$ are well-defined. Due to (22)–(24), $x_i(k)$, $\mu_i(k)$, $\lambda_i(k)$ $\forall i \in \mathcal{V} - u(k)$ are well-defined. To show that $x_i(k)$, $\mu_i(k)$, $\lambda_i(k)$ $\forall i \in u(k)$ are well-defined, note from Lemma 1 that there exists a unique $x' \in \mathbb{R}^n$ that minimizes the uniformly strictly convex and continuously differentiable function $\sum_{i\in u(k)} f_i(x) - x^T \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1))$ over Ω. Also, there exist unique $\mu' \in \mathbb{R}^p$ and $\lambda' \in \mathbb{R}^m$ such that x', μ', and λ' uniquely satisfy the KKT conditions

$$\begin{aligned}
&\sum_{i\in u(k)} \nabla_x \mathcal{L}_i(x', \mu', \lambda') \\
&\quad = \sum_{i\in u(k)} \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)), \\
&x' \in \Omega, \\
&\mu'^{(\ell)} \geq 0, \quad \forall \ell \in \mathbb{P}_p, \\
&\mu'^{(\ell)} g^{(\ell)}(x') = 0, \quad \forall \ell \in \mathbb{P}_p.
\end{aligned}$$

This, along with (25), (26)–(28), (9)–(11), implies that $\forall i \in u(k)$, $x_i(k) = x'$, $\mu_i(k) = \mu'$, $\lambda_i(k) = \lambda'$, which are uniquely defined. Therefore, $x_i(k)$, $\mu_i(k)$, $\lambda_i(k)$ $\forall k \in \mathbb{N}$ $\forall i \in \mathcal{V}$ are well-defined. □

Proposition 3 and its proof suggest that at each iteration $k \in \mathbb{P}$, the gossiping pair $u_1(k)$ and $u_2(k)$ can update their state variables $x_{u_1(k)}(k)$, $x_{u_2(k)}(k)$, $\mu_{u_1(k)}(k)$, $\mu_{u_2(k)}(k)$, $\lambda_{u_1(k)}(k)$, and $\lambda_{u_2(k)}(k)$ by first finding the unique minimizer x' and unique Lagrange multipliers μ' and λ' of the local constrained convex optimization problem

$$\min_{x\in\Omega} \sum_{i\in u(k)} f_i(x) - x^T \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)). \tag{29}$$

Upon solving (29), nodes $u_1(k)$ and $u_2(k)$ set $x_{u_1(k)}(k)$ and $x_{u_2(k)}(k)$ to x', $\mu_{u_1(k)}(k)$ and $\mu_{u_2(k)}(k)$ to μ', and $\lambda_{u_1(k)}(k)$ and $\lambda_{u_2(k)}(k)$ to λ'. Note that to solve (29), one of the two nodes, say, node $u_1(k)$, must send the other node, node $u_2(k)$, its function $f_{u_1(k)}$ at least once. Since $f_{u_1(k)}$ is an infinite-dimensional object (unless $f_{u_1(k)}$ can be analytically expressed), this represents a drawback of our approach. With that said, we point out that for the one-dimensional, unconstrained version of problem (2), this drawback can be eliminated by a bisectioning algorithm [18].

Expressions (19)–(21), (22)–(24), (25), (26)–(28), and (9)–(11) collectively define the following gossip-style, distributed asynchronous iterative algorithm, which we refer to as *Pairwise Equalizing* (PE):

Algorithm 1 (Pairwise Equalizing)**.**
Initialization:
1) Each node $i \in \mathcal{V}$ computes $x_i^* \in \Omega$, $\mu_i^* \in [0,\infty)^p$, and $\lambda_i^* \in \mathbb{R}^m$ by solving (15)–(18).
2) Each node $i \in \mathcal{V}$ creates variables $x_i \in \Omega$, $\lambda_i \in [0,\infty)^p$, and $\mu_i \in \mathbb{R}^m$ and initializes them: $x_i \leftarrow x_i^*$, $\mu_i \leftarrow \mu_i^*$, $\lambda_i \leftarrow \lambda_i^*$.

Operation: At each iteration:
3) A node with one or more one-hop neighbors, say, node i, initiates the iteration and selects a one-hop neighbor, say, node j, to gossip.
4) Nodes i and j select one of two ways to gossip by labeling themselves as either nodes a and b, or nodes b and a, respectively, where $\{a,b\} = \{i,j\}$.
5) If node b does not know f_a, then node a transmits f_a to node b.
6) Node a transmits x_a, μ_a, and λ_a to node b.
7) Node b determines $\hat{x} \in \Omega$, $\hat{\lambda} \in [0,\infty)^p$, and $\hat{\mu} \in \mathbb{R}^m$ by solving:
$\nabla_x \mathcal{L}_a(\hat{x}, \hat{\mu}, \hat{\lambda}) + \nabla_x \mathcal{L}_b(\hat{x}, \hat{\mu}, \hat{\lambda}) = \nabla_x \mathcal{L}_a(x_a, \mu_a, \lambda_a) + \nabla_x \mathcal{L}_b(x_b, \mu_b, \lambda_b)$,
$\hat{x} \in \Omega$,
$\hat{\mu}^{(\ell)} \geq 0$ $\forall \ell \in \mathbb{P}_p$,
$\hat{\mu}^{(\ell)} g^{(\ell)}(\hat{x}) = 0$ $\forall \ell \in \mathbb{P}_p$.
8) Node b updates x_b, μ_b, and λ_b: $x_b \leftarrow \hat{x}$, $\mu_b \leftarrow \hat{\mu}$, $\lambda_b \leftarrow \hat{\lambda}$.
9) Node b transmits x_b, μ_b, and λ_b to node a.
10) Node a updates x_a, μ_a, and λ_a: $x_a \leftarrow x_b$, $\mu_a \leftarrow \mu_b$, $\lambda_a \leftarrow \lambda_b$. ■

PE in Algorithm 1 consists of an initialization part that is executed once and an operation part that is executed iteratively. In Steps 1 and 2, each node $i \in \mathcal{V}$ determines its initial states on its own by solving (15)–(18), or equivalently, solving the constrained convex optimization problem $\min_{x\in\Omega} f_i(x)$ for the unique minimizer and Lagrange multipliers. Step 3 may be performed either deterministically or stochastically. Step 4 offers the gossiping nodes i and j an opportunity to pick one of two ways to gossip, which lead to the same updated values of their state variables but require different communication and computational efforts. This can been seen from Steps 5–10, whereby the node that labels itself as node b has to compute, while the node that labels itself as node a has to communicate more (except that the condition "node b does not know f_a" in Step 5 is false). Clearly, such asymmetric actions help nodes i and j to better utilize their communication and computational resources. Note that Step 5 is a conditional step that does not have to be carried out if node b has already stored f_a in its memory, in which case $2(n+p+m)$ real-number transmissions are needed for each iteration. Also, Step 5, together with Step 6, enables node b to update its state variables in Steps 7 and 8, where Step 7 is equivalent to solving the local optimization problem (29). Finally, Step 9 is needed so that node a can perform Step 10.

IV. Convergence Analysis

As was stated in Section III, PE guarantees the conservation condition (8)–(11) and tries to ensure the dissipation condition (12) by imposing the equalizing condition (26)–(28). In this section, we show that, with sufficiently rich gossiping pattern, PE succeeds in ensuring the dissipation condition (12) and, thus, achieves asymptotic convergence to x^*.

To show that the $x_i(k)$'s asymptotically converge to x^*, i.e., (7) holds, consider a function $V : \Omega^N \times [0,\infty)^{pN} \times \mathbb{R}^{mN} \to \mathbb{R}$, defined as:

$$\begin{aligned} &V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \\ &= \sum_{i\in\mathcal{V}} \mathcal{L}_i(x^*, \mu^*, \lambda^*) - \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) \\ &\quad - \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k))^T (x^* - x_i(k)), \end{aligned} \tag{30}$$

where $\mathbf{x}(k) \in \Omega^N$, $\boldsymbol{\mu}(k) \in [0,\infty)^{pN}$, and $\boldsymbol{\lambda}(k) \in \mathbb{R}^{mN}$ are obtained by stacking the $x_i(k)$'s, $\mu_i(k)$'s, and $\lambda_i(k)$'s. Moreover, for convenience, let $\mathbf{x}^*$, $\boldsymbol{\mu}^*$, and $\boldsymbol{\lambda}^*$ denote the vectors obtained by stacking N copies of x^*, μ^*, and λ^*, respectively. Note from (9), (10), (4), and (5) that indeed $\mathbf{x}(k) \in \Omega^N$ $\forall k \in \mathbb{N}$, $\boldsymbol{\mu}(k) \in [0,\infty)^{pN}$ $\forall k \in \mathbb{N}$, $\mathbf{x}^* \in \Omega^N$, and $\boldsymbol{\mu}^* \in [0,\infty)^{pN}$. Also note from Assumption 1 that the functions $\mathcal{L}_i(x,\mu,\lambda)$ $\forall i \in \mathcal{V}$ are continuously differentiable with respect to x. Hence, V is well-defined and continuous.

The following two lemmas show that with PE, the sequence $(V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)))_{k=0}^\infty$ is nonnegative and non-increasing:

Lemma 2. *Consider the use of PE described in Algorithm 1. Suppose Assumption 1 holds. Then, for any given* $(u(k))_{k=1}^\infty$, $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \geq 0$ $\forall k \in \mathbb{N}$, *where the equality holds if and only if* $\mathbf{x}(k) = \mathbf{x}^*$. *Moreover,*

$$V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \geq \sum_{i\in\mathcal{V}} \gamma_i(\|x^* - x_i(k)\|), \tag{31}$$

where $\gamma_i : [0,\infty) \to [0,\infty)$ *is a continuous and strictly increasing function satisfying* $\gamma_i(0) = 0$, $\lim_{d\to\infty} \gamma_i(d) = \infty$, *and* $f_i(y) - f_i(x) - \nabla f_i(x)^T(y-x) \geq \gamma_i(\|y-x\|)$ $\forall x, y \in \mathbb{R}^n$.

Proof. Let $(u(k))_{k=1}^\infty$ be given and $k \in \mathbb{N}$. Due to (4) and (10), we have $\mu_i^{(\ell)}(k) g^{(\ell)}(x^*) \leq 0$ $\forall \ell \in \mathbb{P}_p$ $\forall i \in \mathcal{V}$. This, along with (4), (6), and (9), implies that $\mathcal{L}_i(x^*, \mu^*, \lambda^*) \geq \mathcal{L}_i(x^*, \mu_i(k), \lambda_i(k))$ $\forall i \in \mathcal{V}$. Hence, $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \geq \sum_{i\in\mathcal{V}} \mathcal{L}_i(x^*, \mu_i(k), \lambda_i(k)) - \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) - \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k))^T(x^* - x_i(k)) = \sum_{i\in\mathcal{V}} \Big(f_i(x^*) - f_i(x_i(k)) - \nabla f_i(x_i(k))^T(x^* - x_i(k)) \Big) + \sum_{\ell=1}^p \mu_i^{(\ell)}(k) \Big(g^{(\ell)}(x^*) - g^{(\ell)}(x_i(k)) - \nabla g^{(\ell)}(x_i(k))^T(x^* - x_i(k)) \Big) + \sum_{\ell=1}^m \lambda_i^{(\ell)}(k) \Big(h^{(\ell)}(x^*) - h^{(\ell)}(x_i(k)) - \nabla h^{(\ell)}(x_i(k))^T(x^* - x_i(k)) \Big)$. For each $i \in \mathcal{V}$, since f_i is uniformly strictly convex, there exists a continuous and strictly increasing function $\gamma_i : [0,\infty) \to [0,\infty)$ satisfying $\gamma_i(0) = 0$ and $\lim_{d\to\infty} \gamma_i(d) = \infty$ such that $f_i(y) - f_i(x) - \nabla f_i(x)^T(y-x) \geq \gamma_i(\|y-x\|)$ $\forall x, y \in \mathbb{R}^n$. Thus, the expression within the first big parentheses is no less than $\gamma_i(\|x^* - x_i(k)\|)$. Moreover, for each $\ell \in \mathbb{P}_p$, since $g^{(\ell)}$ is convex, the expression within the second big parentheses is nonnegative. Furthermore, for each $\ell \in \mathbb{P}_m$, since $h^{(\ell)}$ is affine, the expression within the third big parentheses is zero. It follows from (10) that (31) holds, implying that $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \geq 0$. In addition, due to (4), (6), (9), (11), and (30), $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) = 0$ if $\mathbf{x}(k) = \mathbf{x}^*$. Because of this and (31) and because $\gamma(d) = 0$ if and only if $d = 0$, $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) = 0$ if and only if $\mathbf{x}(k) = \mathbf{x}^*$. ☐

Lemma 3. *Consider the use of PE described in Algorithm 1. Suppose Assumption 1 holds. Then, for any given* $(u(k))_{k=1}^\infty$, $(V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)))_{k=0}^\infty$ *is non-increasing and satisfies*

$$\begin{aligned} &V(\mathbf{x}(k-1), \boldsymbol{\mu}(k-1), \boldsymbol{\lambda}(k-1)) - V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \\ &\geq \sum_{i\in u(k)} \gamma_i(\|x_i(k) - x_i(k-1)\|), \quad \forall k \in \mathbb{P}, \end{aligned} \tag{32}$$

where the γ_i*'s are as in Lemma 2.*

Proof. Let $(u(k))_{k=1}^\infty$ be given and $k \in \mathbb{P}$. From (22)–(24), (25), and (26), $V(\mathbf{x}(k-1), \boldsymbol{\mu}(k-1), \boldsymbol{\lambda}(k-1)) - V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) = \sum_{i\in u(k)} -\mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)) + \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)) x_i(k-1) + \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) - \nabla_x \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) x_i(k) = \sum_{i\in u(k)} \mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) - \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)) - \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1))^T (x_i(k) - x_i(k-1))$. Due to (9), (10), and (11), we have $\mathcal{L}_i(x_i(k), \mu_i(k), \lambda_i(k)) \geq \mathcal{L}_i(x_i(k), \mu_i(k-1), \lambda_i(k-1))$. Hence, $V(\mathbf{x}(k-1), \boldsymbol{\mu}(k-1), \boldsymbol{\lambda}(k-1)) - V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \geq \sum_{i\in u(k)} \mathcal{L}_i(x_i(k), \mu_i(k-1), \lambda_i(k-1)) - \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1)) - \nabla_x \mathcal{L}_i(x_i(k-1), \mu_i(k-1), \lambda_i(k-1))^T (x_i(k) - x_i(k-1)) = \sum_{i\in u(k)} \Big(f_i(x_i(k)) - f_i(x_i(k-1)) - \nabla f_i(x_i(k-1))^T (x_i(k) - x_i(k-1)) \Big) + \sum_{\ell=1}^p \mu_i^{(\ell)}(k-1) \Big(g^{(\ell)}(x_i(k)) - g^{(\ell)}(x_i(k-1)) - \nabla g^{(\ell)}(x_i(k-1))^T (x_i(k) - x_i(k-1)) \Big) + \sum_{\ell=1}^m \lambda_i^{(\ell)}(k-1) \Big(h^{(\ell)}(x_i(k)) - h^{(\ell)}(x_i(k-1)) - \nabla h^{(\ell)}(x_i(k-1))^T (x_i(k) - x_i(k-1)) \Big)$. Similar to the proof of Lemma 2, it can be shown that (32) holds, implying that $(V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)))_{k=0}^\infty$ is non-increasing. ☐

Lemmas 2 and 3 suggest that $V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k))$ may be viewed as a Lyapunov-like function with respect to $\mathbf{x}(k)$, taking nonnegative values in general and the value of zero if and only if $\mathbf{x}(k) = \mathbf{x}^*$. Moreover, the function lies "above" a radially unbounded function $\sum_{i\in\mathcal{V}} \gamma_i(\|x^* - x_i(k)\|)$, with values that are non-increasing as PE executes. This alone, however, is insufficient to claim that the dissipation condition (12) holds. Indeed, to satisfy (12), some restrictions must be placed on the gossiping patterns $(u(k))_{k=1}^\infty$.

To this end, let

$$\mathcal{E}_\infty = \{\{i,j\} : u(k) = \{i,j\} \text{ for infinitely many } k \in \mathbb{P}\},$$

so that a link $\{i,j\}$ is in $\mathcal{E}_\infty$ if and only if nodes i and j gossip with each other infinitely often. Consider the following assumption on the gossiping patterns, which was first proposed in [22]:

Assumption 2. The sequence $(u(k))_{k=1}^\infty$ is such that the graph $(\mathcal{V}, \mathcal{E}_\infty)$ is connected.

The following proposition says that, with gossiping patterns that satisfy Assumption 2, PE does guarantee the dissipation conditions (12):

Proposition 4. *Consider the use of PE described in Algorithm 1. Suppose Assumptions 1 and 2 hold. Then,* (12) *holds.*

Proof. Suppose Assumptions 1 and 2 hold. From Lemmas 2 and 3, $\lim_{k\to\infty} V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) = c$ for some

$c \geq 0$. Thus, $\forall \epsilon' > 0$, $\exists k_1 \in \mathbb{N}$ such that $c \leq V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) < c + \epsilon'$ $\forall k \geq k_1$, implying that $0 \leq V(\mathbf{x}(k_1), \boldsymbol{\mu}(k_1), \boldsymbol{\lambda}(k_1)) - V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) < \epsilon'$ $\forall k \geq k_1$. Due to this and (32), we have $\sum_{k=k_1}^{\infty} \sum_{i \in u(k)} \gamma_i(\|x_i(k) - x_i(k-1)\|) < \epsilon'$, where γ_i $\forall i \in \mathcal{V}$ are as in Lemma 2. Let $\gamma : [0, \infty) \to [0, \infty)$ be defined as $\gamma(x) = \min_{i \in \mathcal{V}} \gamma_i(x)$. Then, γ is a continuous and strictly increasing function satisfying $\gamma(0) = 0$ and $\lim_{d \to \infty} \gamma(d) = \infty$. From (22), we have $\sum_{k=k_1}^{\infty} \sum_{i \in \mathcal{V}} \gamma(\|x_i(k) - x_i(k-1)\|) < \epsilon'$. Hence, $\lim_{k \to \infty} \gamma(\|x_i(k) - x_i(k-1)\|) = 0$ $\forall i \in \mathcal{V}$. Due to the continuity of γ, $\lim_{k \to \infty} \|x_i(k) - x_i(k-1)\| = 0$. Thus, $\lim_{k \to \infty} x_i(k) = s_i$ for some $s_i \in \Omega$. Let $\{\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_q\} = \{s_1, s_2, \ldots, s_N\}$ where $q \in \mathcal{V}$ and $\tilde{s}_i \neq \tilde{s}_j$ $\forall i, j \in \mathbb{P}_q$, $i \neq j$. Now suppose $q \geq 2$. Let $\epsilon > 0$ be such that $\min_{i,j \in \mathbb{P}_q,\, i \neq j} \|\tilde{s}_i - \tilde{s}_j\| = 4\gamma^{-1}(\epsilon)$. Then, $\exists k_2 \in \mathbb{N}$ such that $0 \leq V(\mathbf{x}(k-1), \boldsymbol{\mu}(k-1), \boldsymbol{\lambda}(k-1)) - V(\mathbf{x}(k), \boldsymbol{\mu}(k), \boldsymbol{\lambda}(k)) \leq \epsilon$ $\forall k \geq k_2 + 1$. It follows from (32) that $\|x_i(k) - x_i(k-1)\| < \gamma^{-1}(\epsilon)$ $\forall k \geq k_2 + 1$ $\forall i \in u(k)$. Because of (26) and the triangle inequality,

$$\|x_i(k-1) - x_j(k-1)\| < 2\gamma^{-1}(\epsilon), \quad \forall k \geq k_2 + 1,\ \forall i \in u(k). \tag{33}$$

Also, since $\lim_{k \to \infty} x_i(k) = s_i$ $\forall i \in \mathcal{V}$, $\exists k_3 \in \mathbb{N}$ such that $\|x_i(k) - s_i\| < \gamma^{-1}(\epsilon)$ $\forall k \geq k_3$ $\forall i \in \mathcal{V}$. Let $K = \max\{k_2, k_3\}$. Then, $\|x_i(k) - x_j(k)\| \geq 2\gamma^{-1}(\epsilon)$ $\forall k \geq K$ $\forall i, j \in \mathcal{V}$ with $s_i \neq s_j$. This, along with (33), implies that $\forall i, j \in \mathcal{V}$ with $s_i \neq s_j$, $\{i, j\} \neq u(k)$ $\forall k \geq K + 1$. Hence, there exist $q \geq 2$ disjoint nonempty subsets $U_1, U_2, \ldots, U_q$ of $\mathcal{V}$ with $\cup_{i=1}^{q} U_i = \mathcal{V}$, such that $\forall k \geq K + 1$, $u(k)$ is a subset of one of them. This implies that the graph $(\mathcal{V}, \mathcal{E}_\infty)$ is not connected, i.e., Assumption 2 is violated. Therefore, $q = 1$, i.e., (12) holds. ☐

Built upon Proposition 4, the following theorem shows that the $x_i(k)$'s indeed asymptotically converge to x^* under Assumption 2, solving problem (2):

Theorem 2. *Consider the use of PE described in Algorithm 1. Suppose Assumptions 1 and 2 hold. Then,* (7) *holds.*

Proof. The theorem is an immediate consequence of Theorem 1 and Proposition 4. ☐

Remark 2. We point out that the above analysis establishes only the asymptotic convergence of the $x_i(k)$'s to x^*. It says nothing about the behavior of the Lagrange multipliers $\mu_i(k)$'s and $\lambda_i(k)$'s. Although the primary goal is to achieve (7), it may be of interest to study in future work how these multipliers behave.

V. Conclusion

In this paper, we have introduced and analyzed PE, a gossip algorithm that solves a class of distributed convex optimization problems with identical constraints over networks with time-varying topologies. We have shown that unlike the existing subgradient algorithms, PE does not require stepsizes nor projection to operate. Instead, it strives to satisfy a conservation and dissipation condition induced by the KKT conditions. We have shown that these two conditions can be met by solving, in pairwise fashion, a sequence of suitably defined local optimization problems. Finally, we have shown that under a mild assumption on the gossiping pattern, PE is asymptotically convergent to the unknown minimizer, solving the given problem.

References

[1] B. Johansson, M. Rabi, and M. Johansson, "A simple peer-to-peer algorithm for distributed optimization in sensor networks," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 4705–4710.

[2] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. IEEE Conference on Decision and Control*, Cancun, Mexico, 2008, pp. 4185–4190.

[3] I. Lobel and A. Ozdaglar, "Convergence analysis of distributed subgradient methods over random networks," in *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, 2008, pp. 353–360.

[4] A. Nedić and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[5] A. Nedić, D. P. Bertsekas, and V. S. Borkar, "Distributed asynchronous incremental subgradient methods," in *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, D. Butnariu, Y. Censor, and S. Reich, Eds. Amsterdam, Holland: Elsevier, 2001, pp. 381–407.

[6] A. Nedić and D. P. Bertsekas, "Convergence rate of incremental subgradient algorithms," in *Stochastic Optimization: Algorithms and Applications*, S. P. Uryasev and P. M. Pardalos, Eds. Norwell, MA: Kluwer Academic Publishers, 2001, pp. 223–264.

[7] A. Nedić and A. Ozdaglar, "On the rate of convergence of distributed subgradient methods for multi-agent optimization," in *Proc. IEEE Conference on Decision and Control*, New Orleans, LA, 2007, pp. 4711–4716.

[8] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proc. IEEE Conference on Decision and Control*, Cancun, Mexico, 2008, pp. 4177–4184.

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

[10] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[11] M. G. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *Proc. International Symposium on Information Processing in Sensor Networks*, Berkeley, CA, 2004, pp. 20–27.

[12] ——, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798–808, 2005.

[13] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Stochastic incremental gradient descent for estimation in sensor networks," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 2007, pp. 582–586.

[14] ——, "Incremental stochastic subgradient algorithms for convex optimization," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 691–717, 2009.

[15] ——, "Asynchronous gossip algorithms for stochastic optimization," in *Proc. IEEE Conference on Decision and Control*, Shanghai, China, 2009, pp. 3581–3586.

[16] ——, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[17] S.-H. Son, M. Chiang, S. R. Kulkarni, and S. C. Schwartz, "The value of clustering in distributed estimation for sensor networks," in *Proc. International Conference on Wireless Networks, Communications and Mobile Computing*, Maui, HI, 2005, pp. 969–974.

[18] J. Lu, C. Y. Tang, P. R. Regier, and T. D. Bow, "Gossip algorithms for convex consensus optimization over networks," to appear in *IEEE Transactions on Automatic Control*, vol. 56, no. 12, 2011.

[19] J. Lu, P. R. Regier, and C. Y. Tang, "Control of distributed convex optimization," in *Proc. IEEE Conference on Decision and Control*, Atlanta, GA, 2010, pp. 489–495.

[20] E. K. P. Chong and S. H. Zak, *An Introduction To Optimization*. New York, NY: John Wiley & Sons, 1996.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[22] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1984.