

Proximal minimization based distributed convex optimization

Kostas Margellos, Alessandro Falsone, Simone Garatti and Maria Prandini

Abstract—We provide a novel iterative algorithm for distributed convex optimization over time-varying multi-agent networks, in the presence of heterogeneous agent constraints. We adopt a proximal minimization perspective and show that this set-up allows us to bypass the difficulties of existing algorithms while simplifying the underlying mathematical analysis. At every iteration each agent makes a tentative decision by solving a local optimization program, and then communicates this decision with neighboring agents. We show that following this scheme agents reach consensus on a common decision vector, and in particular that this vector is an optimizer of the centralized problem.

I. INTRODUCTION

Optimization in multi-agent networks has attracted significant attention in the control and signal processing literature, due to its applicability in different domains like power systems [1], [2], wireless networks [3], [4], robotics [5], etc. Typically, agents solve a local decision making problem, communicate their decisions with other agents and repeat the process on the basis of the new information received. The main objective of this cooperative set-up is for agents to reach consensus and agree on a common decision that optimizes a certain performance criterion for the overall multi-agent system.

This set-up allows each agent to keep some of its own data private and enables distributed computation, which may lead to computational savings compared to centralized paradigms. Based on the underlying communication structure, two main optimization architectures can be distinguished, namely, decentralized and distributed optimization, according to the terminology of [6]. In the former, computation is distributed across agents but there is a centralized communication infrastructure (e.g., a central authority) that collects information from each agent and transmits updates to all of them. On the other hand, in distributed optimization not only computation but also communication is distributed, and agents exchange information only with neighboring agents.

Most literature builds on the seminal work of [6–8] (see also [9], [10] and references therein for a more recent problem exposition), where a wide range of optimization problems is considered, mostly from a decentralized perspective, using techniques based on gradient descent, dual

decomposition, and the method of multipliers. The recent work [11] deals with similar problems but from a mean field game theoretic perspective, where agents' decisions converge to a Nash equilibrium, not necessarily an optimizer, as the number of agents tends to infinity. Distributed consensus and/or optimization problems, but in the absence of constraints in the underlying optimization programs, have been considered in [12–22]. In most of these references the underlying network is allowed to be time-varying. In the presence of constraints and following a distributed paradigm, the authors of [10], [23–25] adopt Newton based or gradient/subgradient based approaches and show asymptotic agreement of the agents' solutions to an optimizer of the centralized problem. In [26] a distributed alternating direction method of multipliers approach is adopted and its convergence properties are analyzed, whereas in [27], [28] an approach based on constraint exchange is adopted. In these cases, however, the underlying network is time-invariant.

In a time-varying environment, [29–31] follow a projected gradient/subgradient methodology to solve distributed convex optimization problems in the presence of constraints. In [29], however, the particular case where the agents' constraint sets are all identical is considered. As a result, the computational complexity of each agents' local optimization program is the same with the one of the centralized problem. In [30] a more general case is considered, again using a projected subgradient methodology. A similar problem is considered in [31] but the objective functions are required to be differentiable with Lipschitz continuous gradients.

In this paper we deal with distributed convex optimization problems over time-varying networks, under a possibly different constraint set per agent. Our approach is most closely related to the work of [29–31], but we adopt a proximal minimization instead of a gradient/subgradient based perspective. We provide a detailed convergence and optimality analysis of the proposed algorithm, extending the results of [29], [30] (compare our results with the proof of Lemma 5.4 in [30]), while relaxing the differentiability assumptions of [31].

The paper is structured as follows. In Section II we formulate the proposed distributed algorithm based on proximal minimization and state the underlying assumptions. In Section III we provide some preparatory results and establish useful relations regarding the agents' local solutions, whereas in Section IV we show convergence and optimality of the developed algorithm. Finally, Section V concludes the paper and provides some directions for future work. All proofs have been omitted in the interest of space, but can be found in [32].

Research was supported by the European Commission, H2020, under the project UnCoVerCPS, grant number 643921.

Kostas Margellos is with the Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ, Oxford, United Kingdom, e-mail: kostas.margellos@eng.ox.ac.uk

Alessandro Falsone, Simone Garatti and Maria Prandini are with the Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, e-mail: {alessandro.falsone, simone.garatti, maria.prandini}@polimi.it

Algorithm 1 Distributed, optimal consensus

- 1: **Initialization**
 - 2: $k = 0$.
 - 3: Consider $x_i(0) \in X_i$, for all $i = 1, \dots, m$.
 - 4: **For** $i = 1, \dots, m$ **repeat until convergence**
 - 5: $z_i(k) = \sum_{j=1}^m a_j^i(k) x_j(k)$.
 - 6: $x_i(k+1) = \arg \min_{x_i \in X_i} f_i(x_i) + \frac{1}{2c(k)} \|z_i(k) - x_i\|^2$.
 - 7: $k \leftarrow k + 1$.
-

II. DISTRIBUTED CONSTRAINED CONVEX OPTIMIZATION

A. Problem set-up

We consider a time-varying network of m agents that communicate to cooperatively solve an optimization problem of the form

$$\begin{aligned} \mathcal{P} : \quad & \min_{x \in \mathbb{R}^n} \sum_{i=1}^m f_i(x) & (1) \\ & \text{subject to } x \in \bigcap_{i=1}^m X_i, \end{aligned}$$

where $x \in \mathbb{R}^n$ represents a vector of n decision variables. For each $i = 1, \dots, m$, $f_i(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function of agent i , whereas $X_i \subseteq \mathbb{R}^n$ is its constraint set. Since most of the subsequent results are based on $f_i(\cdot)$ and X_i being convex, we formalize it in the following assumption.

Assumption 1: [Convexity] For each $i = 1, \dots, m$, the function $f_i(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ and the set $X_i \subseteq \mathbb{R}^n$ are convex.

Problem \mathcal{P} cannot be solved in a centralized fashion if $f_i(\cdot)$ and X_i represent private information, available only to agent i . Moreover, even in the case where all necessary information (objective functions and constraint sets) was available to all agents, imposing all the constraints in one shot, by encoding $\bigcap_{i=1}^m X_i$, may result in a program which is too intensive computationally. To account for information privacy, and simultaneously alleviate the computational effort, we follow a distributed, iterative approach, where each agent i solves an appropriate, local optimization problem and exchanges information with other agents based on the outcome of this optimization. We will show that under certain structural and communication assumptions agents reach consensus to an optimal solution of \mathcal{P} . The basic steps of the proposed approach are summarized in Algorithm 1, whose interpretation is as follows.

Initially, each agent i , $i = 1, \dots, m$, starts with some arbitrary $x_i(0)$, which constitutes its estimate of what a minimizer of \mathcal{P} might be (step 3, Algorithm 1). This estimate belongs to the local constraint set X_i of agent i , but not necessarily to $\bigcap_{i=1}^m X_i$. One choice for $x_i(0)$ is e.g., $x_i(0) \in \arg \min_{x_i \in X_i} f_i(x_i)$. At iteration k , each agent i constructs a weighted average $z_i(k)$ of those solutions $x_j(k)$, $j = 1, \dots, m$, communicated by the other other agents and its local one (step 5, Algorithm 1). Coefficient $a_j^i(k) \in [0, 1]$ indicates how agent i weights the solution received by agent j at iteration k , and $a_j^i(k) = 0$ encodes the fact that agent i does not receive any information from agent j at iteration k .

Agent i solves then a local minimization problem, seeking the optimal solution within X_i that minimizes a performance criterion, which is defined as a linear combination of the local objective function $f_i(x_i)$ and a quadratic term penalizing the difference from $z_i(k)$ (step 6, Algorithm 1). The relative importance of these two terms is dictated by $c(k) \in \mathbb{R}_+$. Note that, under Assumption 1 and due to the presence of the quadratic penalty term, the resulting problem is strictly convex with respect to x_i , and hence admits a unique solution.

B. Further structural assumptions and communication requirements

We impose some additional assumptions on the structure of problem \mathcal{P} in (1) and the communication set-up that is considered in this paper. These assumptions are crucial for the analysis of Section IV.

Assumption 2: [Compactness and Continuity] For each $i = 1, \dots, m$, $X_i \subseteq \mathbb{R}^n$ is compact. Moreover, $f_i(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz continuous on X_i with Lipschitz constant $L_i \in \mathbb{R}_+$, i.e., for all $i = 1, \dots, m$,

$$|f_i(x) - f_i(y)| \leq L_i \|x - y\|, \text{ for all } x, y \in X_i. \quad (2)$$

Assumption 3: [Interior point] The feasibility region $\bigcap_{i=1}^m X_i$ of \mathcal{P} has a non-empty interior, i.e., there exists $\bar{x} \in \bigcap_{i=1}^m X_i$ and $\rho \in \mathbb{R}_+$ such that $\{x \in \mathbb{R}^n : \|x - \bar{x}\| < \rho\} \subset \bigcap_{i=1}^m X_i$.

Note that due to the compactness condition of Assumption 2, $\text{co}(\bigcup_{i=1}^m X_i)$ is also compact, with $\text{co}(\cdot)$ denoting the convex hull of its argument. Let then $D \in \mathbb{R}_+$ be such that $\|x\| \leq D$ for all $x \in \text{co}(\bigcup_{i=1}^m X_i)$. Moreover, due to compactness and Assumption 3, by the Weierstrass' theorem (Proposition A.8, p. 625 in [6]), \mathcal{P} admits at least one optimal solution. Therefore, if we denote by $X^* \subseteq \bigcap_{i=1}^m X_i$ the set of optimizers of \mathcal{P} , then X^* is non-empty. Notice also that $f_i(\cdot)$, $i = 1, \dots, m$, is continuous due to the convexity condition of Assumption 1; the addition of Assumption 2 is to require Lipschitz continuity. Note that $f_i(\cdot)$, $i = 1, \dots, m$, is not required to be differentiable.

We impose the following assumption on the coefficients $\{c(k)\}_{k \geq 0}$, that appear in step 6 of Algorithm 1.

Assumption 4: [Coefficient $\{c(k)\}_{k \geq 0}$] Assume that for all $k \geq 0$, $c(k) \in \mathbb{R}_+$ and $\{c(k)\}_{k \geq 0}$ is a non-increasing sequence, i.e., $c(k) \leq c(r)$ for all $k \geq r$, with $r \geq 0$. Moreover,

- 1) $\sum_{k=0}^{\infty} c(k) = \infty$,
- 2) $\sum_{k=0}^{\infty} c(k)^2 < \infty$.

In standard proximal minimization [6] convergence is highly dependent on the appropriate choice of $c(k)$. Assumption 4 is in fact needed to guarantee convergence of Algorithm 1 in Section IV-C. A direct consequence of the last part of Assumption 4 is that $\lim_{k \rightarrow \infty} c(k) = 0$. One choice for $\{c(k)\}_{k \geq 0}$ that satisfies the conditions of Assumption 4 is to select it from the class of generalized harmonic series [33]. Note that Assumption 4 is in a sense analogous to the conditions that the authors of [29], [30] impose on the step-size of their subgradient algorithm.

In line with [7], [8], [19] we impose the following assumptions on the information exchange between the agents.

Assumption 5: [Weight coefficients] There exists $\eta \in (0, 1)$ such that for all $i, j \in \{1, \dots, m\}$ and all $k \geq 0$, $a_i^i(k) \geq \eta$, and $a_j^i(k) > 0$ implies that $a_j^i(k) \geq \eta$. Moreover, for all $k \geq 0$,

- 1) $\sum_{j=1}^m a_j^i(k) = 1$ for all $i = 1, \dots, m$,
- 2) $\sum_{i=1}^m a_j^i(k) = 1$ for all $j = 1, \dots, m$.

The interpretation of having a uniform lower bound η , independent of k , for coefficients $a_j^i(k)$ in Assumption 5 is that it ensures that each agent is mixing information received by other agents at a non-diminishing rate in time [29]. Moreover, point 1 in Assumption 5 ensures that this mixing is a convex combination of the other agent estimates, assigning a non-zero weight to its local one due to the fact that $a_i^i(k) \geq \eta$, while point 2 ensures that agents influence each other equally in the long run of the algorithm (see [29] for further elaboration on this issue).

For each $k \geq 0$ the information exchange between the m agents can be represented by a directed graph (V, E_k) , where the nodes $V = \{1, \dots, m\}$ are the agents and the set E_k of directed edges is given by

$$E_k = \{(j, i) : a_j^i(k) > 0\}, \quad (3)$$

i.e., each edge implies that at time k agent i receives information, namely estimate $x_j(k)$ from agent j . If $a_j^i(k) > 0$ we also say that j is a neighboring agent of i at time k . Under this set-up, Algorithm 1 is fully distributed, because at iteration k each agent $i = 1, \dots, m$ receives information only from neighboring agents. This information exchange is time-varying, since the weighting coefficients depend on the iteration index k , and may be occasionally the empty set.

Let $E_\infty = \{(j, i) : (j, i) \in E_k \text{ for infinitely many } k\}$ denote the set of edges (j, i) that represent agent pairs that communicate directly infinitely often. We then impose the following connectivity and communication assumption.

Assumption 6: [Connectivity and Communication] The graph (V, E_∞) is strongly connected, i.e., for any two nodes there exists a path of directed edges that connects them. Moreover, there exists $T \geq 1$ such that for every $(j, i) \in E_\infty$, agent i receives information from a neighboring agent j at least once every consecutive T iterations.

Assumption 6 guarantees that all pairs of agents mix their information infinitely often via some network path (not necessarily with one-line communication), and the intercommunication interval is bounded. For a more formal exposition and further details on the interpretation of the imposed network structure the reader is referred to [18], [29].

Assumptions 5 and 6 are identical to Assumptions 2-5 in [29] (the same assumptions are also imposed in [30]). Note that these are rather standard for distributed consensus problems; for further relaxations the reader is referred to [19], [34].

The proposed iterative methodology resembles the structure of proximal minimization for constrained convex op-

timization [6] (Chapter 3.4.3). The difference, however, is that our set-up is distributed and the quadratic term in step 6 does not penalize the deviation of x_i from the previous iterate $x_i(k)$, but from an appropriately weighted average $z_i(k)$. Our approach has an intuitive economic interpretation; at every iteration k we penalize a consensus residual proxy by the time-varying coefficient $1/(2c(k))$, which, due to Assumption 4, progressively increases. This can be thought of as a pricing settling mechanism, where the more we delay to achieve consensus the higher the price is.

In the case where $a_j^i(k) = 1/m$ for all $i, j = 1, \dots, m$, for all $k \geq 0$, that corresponds to a decentralized control paradigm, the solution of our proximal minimization approach coincides with the one obtained when the alternating direction of multipliers [6], [9], is applied to this problem (see eq. (4.72)-(4.74), p. 254 in [6]). In the latter the quadratic penalty term is not added to the local objective function as in step 6 of Algorithm 1, but to the Lagrangian function of an equivalent problem, and the coefficient $c(k)$ is an arbitrary constant independent of k ; however, a dual-update step is required. Formal connections between penalty methods and the method of multipliers have been established in [35].

III. PREPARATORY RESULTS

In this section several relations between the difference of the agent estimates from certain average quantities are stated. At the end of the section we provide a relation that is fundamental for the analysis of Section IV.

For all $k \geq 0$, let

$$v(k) = \frac{1}{m} \sum_{i=1}^m x_i(k), \text{ for all } k \geq 0. \quad (4)$$

$$\bar{v}(k) = \frac{\epsilon(k)}{\epsilon(k) + \rho} \bar{x} + \frac{\rho}{\epsilon(k) + \rho} v(k), \quad (5)$$

where $\epsilon(k) = \sum_{i=1}^m \text{dist}(v(k), X_i)$, and $\bar{x} \in \mathbb{R}^n$, $\rho \in \mathbb{R}_+$ are as in Assumption 3. It is shown in Lemma 2 of [29] that $\bar{v}(k) \in \bigcap_{i=1}^m X_i$. Note that unlike $x_i(k)$ and $v(k)$, which do not necessarily belong to $\bigcap_{i=1}^m X_i$, for $\bar{v}(k)$ this is always the case, thus providing a feasible solution of \mathcal{P} .

For each $i = 1, \dots, m$, denote by

$$e_i(k+1) = x_i(k+1) - z_i(k), \text{ for all } k \geq 0, \quad (6)$$

the error between the values computed at steps 5 and 6 of Algorithm 1, i.e., the difference of the weighted average $z_i(k)$ computed by agent i at time k from its local update $x_i(k+1)$.

A. Error relations

We provide some intermediate results that form the basis of the relation of the next subsection.

Lemma 1: Consider the compactness condition of Assumption 2 and Assumption 3. For all $k \geq 0$,

$$\sum_{i=1}^m \|x_i(k) - \bar{v}(k)\| \leq \mu \sum_{i=1}^m \|x_i(k) - v(k)\|, \quad (7)$$

where $\mu = (2/\rho)mD + 1$.

The proof of Lemma 1 is given in [32]. From step 5 of Algorithm 1 we have that for all $k \geq 0$, for all $i = 1, \dots, m$,

$$\begin{aligned} x_i(k+1) &= \sum_{j=1}^m a_j^i(k)x_j(k) + x_i(k+1) - z_i(k) \\ &= \sum_{j=1}^m a_j^i(k)x_j(k) + e_i(k+1), \end{aligned} \quad (8)$$

where the last equality follows from (6).

Following [18], for each $k \geq 0$ consider a matrix $A(k) \in \mathbb{R}_+^{m \times m}$ such that $a_j^i(k)$ is the j -th element of its i -th column. For all k, s with $k \geq s$, let $\Phi(k, s) = A(s)A(s+1) \dots A(k-1)A(k)$, with $\Phi(k, k) = A(k)$ for all $k \geq 0$. Denote by $[\Phi(k, s)]_j^i$ element j of column i of $\Phi(k, s)$. It is then shown in [18] that, under Assumption 5, $\Phi(k, s)$ is doubly stochastic, i.e., both its rows and columns sum up to one. Similarly to [18], by propagating (8) in time, it can be shown that for all $k > s$ (the inequality is strict for convenience of the subsequent derivations), for all $i = 1, \dots, m$,

$$\begin{aligned} x_i(k+1) &= \sum_{j=1}^m [\Phi(k, s)]_j^i x_j(s) \\ &+ \sum_{r=s}^{k-1} \sum_{j=1}^m [\Phi(k, r+1)]_j^i e_j(r+1) + e_i(k+1). \end{aligned} \quad (9)$$

For all $k > s$, the last statement, together with (19) and the fact that $\Phi(k, s)$ is a doubly stochastic matrix, leads to

$$\begin{aligned} v(k+1) &= \frac{1}{m} \sum_{j=1}^m x_j(s) \\ &+ \frac{1}{m} \sum_{r=s}^{k-1} \sum_{j=1}^m e_j(r+1) + \frac{1}{m} \sum_{j=1}^m e_i(k+1). \end{aligned} \quad (10)$$

We then have the following lemma, which relates $\|x_i(k+1) - v(k+1)\|$ to $\|e_i(k+1)\|$, $i = 1, \dots, m$.

Lemma 2: Consider Assumptions 5 and 6. For all k, s with $s \geq 0$, $k > s$, and for all $i = 1, \dots, m$,

$$\begin{aligned} \|x_i(k+1) - v(k+1)\| &\leq \lambda q^{k-s} \sum_{j=1}^m \|x_j(s)\| \\ &+ \sum_{r=s}^{k-1} \lambda q^{k-r-1} \sum_{j=1}^m \|e_j(r+1)\| \\ &+ \|e_i(k+1)\| + \frac{1}{m} \sum_{j=1}^m \|e_j(k+1)\|, \end{aligned} \quad (11)$$

where $\lambda = 2(1 + \eta^{-(m-1)T}) / (1 - \eta^{(m-1)T}) \in \mathbb{R}_+$ and $q = (1 - \eta^{(m-1)T})^{1/(m-1)T} \in (0, 1)$.

The proof of Lemma 2 follows from the proof of Lemma 4 in [36]; it can be found in [32].

B. A summability relation

Let $N \in \mathbb{N}_+$ and consider the term

$$2\bar{L} \sum_{k=1}^N c(k) \sum_{i=1}^m \|x_i(k+1) - \bar{v}(k+1)\|, \quad (12)$$

where $\bar{L} = \max_{i=1, \dots, m} L_i$ with L_i defined according to (2). We will show that (12) has an interesting relation with $\sum_{k=1}^N \sum_{i=1}^m \|e_i(k+1)\|^2$, which is used extensively in establishing certain summability results in the next section.

Consider Lemma 1 with $k+1$ in place of k and Lemma 2, summing both sides of (11) with respect to $i = 1, \dots, m$ and setting $s = 0$. After some algebraic manipulations and index changes, we have that

$$\begin{aligned} &2\bar{L} \sum_{k=1}^N c(k) \sum_{i=1}^m \|x_i(k+1) - \bar{v}(k+1)\| \\ &\leq 2m\mu\lambda\bar{L} \sum_{k=1}^N c(k)q^k \sum_{i=1}^m \|x_i(0)\| \\ &+ 2m\mu\lambda\bar{L} \sum_{k=1}^N \sum_{r=0}^{k-1} c(k)q^{k-r-1} \sum_{i=1}^m \|e_i(r+1)\| \\ &+ 4\mu\bar{L} \sum_{k=1}^N c(k) \sum_{i=1}^m \|e_i(k+1)\|, \end{aligned} \quad (13)$$

We then have the following lemma, which is the main result of this section.

Lemma 3: Consider Assumptions 2-6. Fix any $\alpha_1 \in (0, 1)$. We then have that for any $N \in \mathbb{N}_+$,

$$\begin{aligned} &2\bar{L} \sum_{k=1}^N c(k) \sum_{i=1}^m \|x_i(k+1) - \bar{v}(k+1)\| \\ &< \alpha_1 \sum_{k=1}^N \sum_{i=1}^m \|e_i(k+1)\|^2 + \alpha_2 \sum_{k=1}^N c(k)^2 + \alpha_3, \end{aligned} \quad (14)$$

where

$$\begin{aligned} \alpha_2 &= \frac{2}{\alpha_1} m\mu^2 \bar{L}^2 \left(m^2 \lambda^2 \frac{1}{(1-q)^2} + 4 \right), \\ \alpha_3 &= 2m^3 \mu^2 \lambda^2 \bar{L}^2 c(0)^2 \frac{1}{\alpha_1 (1-q)^2} \\ &+ 2m^2 \mu \lambda \bar{L} D c(1) \frac{q}{1-q} + 2\alpha_1 m D^2. \end{aligned} \quad (15)$$

The proof of Lemma 3 involves bounding each term in (13) by affine functions of $\sum_{k=1}^N \sum_{i=1}^m \|e_i(k+1)\|^2$ and $\sum_{k=1}^N c(k)^2$, thus resulting in (14). See [32] for details.

IV. CONVERGENCE, CONSENSUS AND OPTIMALITY

In this section we deal with the convergence properties of Algorithm 1, and show that agents reach consensus to some minimizer of \mathcal{P} . Here for the sake of brevity, we just state the results and briefly outline the derivation scheme. The interested reader is referred again to [32] for detailed proofs.

A. Error convergence

We use the following result, which is proven in Lemma 4.1 in [6]. If $y^* = \arg \min_{y \in Y} J_1(y) + J_2(y)$ (assuming uniqueness of the minimizer), where $Y \subseteq \mathbb{R}^n$ is a convex set, $J_1(\cdot), J_2(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex functions and $J_2(\cdot)$ is continuously differentiable, then $y^* = \arg \min_{y \in Y} J_1(y) + \nabla J_2(y^*)^\top y$, where $\nabla J_2(y^*)$ is the gradient of $J_2(y)$ with respect to y , evaluated at y^* . Note that the proof of this statement depends strongly on the convexity requirement. Our analysis deviates from the approach adopted in [29], [30], it is motivated by the proof of the alternating direction method of multipliers (Proposition 4.2 in [6], Appendix A of [9]) and relies on our proximal minimization perspective.

Consider step 6 of Algorithm 1. Under Assumption 1 and since $(1/(2c(k)))\|z_i(k) - x_i\|^2$ is continuously differentiable with respect to x_i , applying the previous result to this problem with x_i, X_i in place of y, Y , respectively, $f_i(x_i)$ in place of $J_1(y)$ and $(1/(2c(k)))\|z_i(k) - x_i\|^2$ in place of $J_2(y)$, we have that

$$x_i(k+1) = \arg \min_{x_i \in X_i} f_i(x_i) - \frac{1}{c(k)}(z_i(k) - x_i(k+1))^\top x_i, \quad (16)$$

where in the second term of (16), $-(1/c(k))(z_i(k) - x_i(k+1))$, is the differential of $(1/(2c(k)))\|z_i(k) - x_i\|^2$ with respect to x_i , evaluated at $x_i(k+1)$. Based on (16), we then have the following lemma, which provides a useful relation between the consecutive algorithm iterates $x_i(k+1)$ and $x_i(k)$, and forms the basis for the subsequent results.

Lemma 4: Consider Assumptions 1-3 and Assumption 5. We then have that for any $k \in \mathbb{N}_+$, for any $x^* \in X^*$,

$$\begin{aligned} & 2c(k) \sum_{i=1}^m f_i(\bar{v}(k+1)) + \sum_{i=1}^m \|e_i(k+1)\|^2 \\ & + \sum_{i=1}^m \|x_i(k+1) - x^*\|^2 \\ & \leq 2c(k) \sum_{i=1}^m f_i(x^*) + \sum_{i=1}^m \|x_i(k) - x^*\|^2 \\ & + 2\bar{L}c(k) \sum_{i=1}^m \|x_i(k+1) - \bar{v}(k+1)\|. \end{aligned} \quad (17)$$

By manipulating (17) as in [32], we are able to show the following proposition.

Proposition 1: Consider Assumptions 1-6. We have that

- 1) $\sum_{k=1}^{\infty} \sum_{i=1}^m \|e_i(k)\|^2 < \infty$,
- 2) $\lim_{k \rightarrow \infty} \|e_i(k)\| = 0$, for all $i = 1, \dots, m$.

B. Consensus

Under the structural assumptions and the communication set-up imposed in Section II-B, agents reach consensus to a common decision vector. This is formally stated in the following proposition.

Proposition 2 (Consensus): Consider Assumptions 1-6. We have that

$$\lim_{k \rightarrow \infty} \|x_i(k) - v(k)\| = 0, \text{ for all } i = 1, \dots, m, \quad (18)$$

where, we recall,

$$v(k) = \frac{1}{m} \sum_{i=1}^m x_i(k), \text{ for all } k \geq 0. \quad (19)$$

Proposition 2 shows that consensus is reached, and for each $i = 1, \dots, m$, as $k \rightarrow \infty$, $x_i(k)$ converges to the arithmetic average $v(k)$ of the agents' estimates, as this is given by (19). Based on the developments of Section III, the proof of Proposition 2 is similar to the proof of Lemma 4 in [36]; it can be found in [32].

C. Convergence and optimality

We show that Algorithm 1 converges, and in particular that the sequence $\{\|x_i(k) - x^*\|\}_{k \geq 0}$ is convergent for any minimizer $x^* \in X^*$ of \mathcal{P} , for all $i = 1, \dots, m$.

To achieve this, notice that by the first part of Proposition 1 (under Assumptions 1-6), $\sum_{k=1}^{\infty} \sum_{i=1}^m \|e_i(k)\|^2 < \infty$. Letting then $N \rightarrow \infty$ in (14) leads to the following summability result, which states that

$$2\bar{L} \sum_{k=1}^{\infty} c(k) \sum_{i=1}^m \|x_i(k+1) - \bar{v}(k+1)\| < \infty. \quad (20)$$

The last statement enables us to show the following convergence result [32].

Theorem 1: Consider Assumptions 1-6. We have that for all $x^* \in X^*$, the sequence $\{\|x_i(k) - x^*\|\}_{k \geq 0}$ is convergent for all $i = 1, \dots, m$.

Theorem 1 shows that Algorithm 1 converges, however, the iterates $\|x_i(k) - x^*\|$ do not necessarily converge to zero, but to some constant. The next theorem shows that for some minimizer $x^* \in X^*$ of \mathcal{P} the iterates indeed tend to zero, i.e., $\lim_{k \rightarrow \infty} \|x_i(k) - x^*\| = 0$ for some $x^* \in X^*$, for all $i = 1, \dots, m$. The proof can be found in [32].

Theorem 2 (Optimality): Consider Assumptions 1-6. We have that, for some $x^* \in X^*$,

$$\lim_{k \rightarrow \infty} \|x_i(k) - x^*\| = 0, \text{ for all } i = 1, \dots, m. \quad (21)$$

Note that a direct byproduct of Proposition 2, Theorem 2 and Lemma 1, is that the consensus vector is some minimizer of \mathcal{P} , since, for some $x^* \in X^*$, $\lim_{k \rightarrow \infty} \|x_i(k) - x^*\| = \lim_{k \rightarrow \infty} \|v(k) - x^*\| = \lim_{k \rightarrow \infty} \|\bar{v}(k) - x^*\|$, for all $i = 1, \dots, m$. Moreover, by the definition of $\epsilon(k)$ (see below (5)) we also have that $\lim_{k \rightarrow \infty} \epsilon(k) = 0$. The preceding statements imply that the sequences $\{v(k)\}_{k \geq 0}$, $\{\bar{v}(k)\}_{k \geq 0}$ and $\{x_i(k)\}_{k \geq 0}$, $i = 1, \dots, m$, converge to the same element of X^* .

V. CONCLUSION

In this paper we provided a novel algorithm for distributed convex optimization over time-varying multi-agent networks, in the presence of heterogeneous agent constraints. A proximal minimization approach was adopted and an

iterative distributed algorithm was developed. Convergence and optimality of this distributed scheme was shown.

Current work concentrates on three main directions: 1) Investigating the convergence rate properties of the developed algorithm, 2) Developing rolling horizon implementations, extending the work of [37] to the case where constraints are also present, and 3) Extending our results to the case where a “budget” type coupling equality constraint is added to the problem (common in resource allocation problems [38]), and investigate the pricing implications of a distributed scheme for such set-up.

From an application point of view, the main focus is on applying the proposed algorithm to the problem of energy efficient control of a building network.

REFERENCES

- [1] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, “Distributed reactive power feedback control for voltage regulation and loss minimization,” *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 966–981, 2015.
- [2] Y. Zhang and G. Giannakis, “Distributed stochastic market clearing with high-penetration wind power and large-scale demand response,” *IEEE Transactions on Power Systems*, to appear, pp. 1–12, 2015.
- [3] G. Mateos and G. Giannakis, “Distributed recursive least-squares: Stability and performance analysis,” *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3740–3754, 2012.
- [4] B. Baingana, G. Mateos, and G. Giannakis, “Proximal-gradient algorithms for tracking cascades over social networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 563–575, 2014.
- [5] S. Martinez, F. Bullo, J. Cortez, and E. Frazzoli, “On synchronous robotic networks - Part I: Models, tasks, and complexity,” *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2199–2213, 2007.
- [6] D. Bertsekas and J. Tsitsiklis, *Parallel and distributed computation: Numerical methods*. Athena Scientific (republished in 1997), 1989.
- [7] J. Tsitsiklis, *Problems in decentralized decision making and computation*. Ph.D. Dissertation, MIT, Cambridge, MA, 1984.
- [8] J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.
- [9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “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, 2010.
- [10] I. Necoara, V. Nedelcu, and I. Dumitrache, “Parallel and distributed optimization methods for estimation and control in networks,” *Journal of Process Control*, vol. 21, no. 5, pp. 756–766, 2011.
- [11] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, “Decentralized convergence to Nash equilibria in constrained mean field control,” *IEEE Transactions on Automatic Control (to appear)*, pp. 1–15, 2016.
- [12] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet, “Novel type of phase transitions in a system of self-driven particles,” *Physical Review Letters*, vol. 75, no. 6, pp. 1226–1229, 1995.
- [13] A. Jadbabaie, J. Lin, and S. Morse, “Coordination of groups of mobile autonomous agents using nearest neighbor rules,” *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [14] R. Olfati-Saber and R. Murray, “Consensus problems in networks of agents with switching topology and time delays,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [15] M. Cao, D. Spielman, and S. Morse, “A lower bound on convergence of a distributed network convergence algorithm,” in *Proceedings of IEEE Control and Decision Conference*, pp. 2356–2361, 2005.
- [16] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
- [17] B. Johansson, M. Rabi, and M. Johansson, “A simple peer-to-peer algorithm for distributed optimization in sensor networks,” in *Proceedings of IEEE Conference on Decision and Control*, pp. 4705–4710, 2007.
- [18] 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.
- [19] A. Olshevsky and J. Tsitsiklis, “Convergence speed in distributed convergence and averaging,” *SIAM Review*, vol. 53, no. 4, pp. 747–772, 2011.
- [20] H. Ishii and R. Tempo, “The PageRank problem, multi-agent consensus, and web aggregation: A systems and control viewpoint,” *IEEE Control Systems Magazine*, vol. 34, no. 3, pp. 34–53, 2014.
- [21] A. Nedic and A. Olshevsky, “Distributed optimization over time-varying directed graphs,” *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.
- [22] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, “Newton-Raphson consensus for distributed convex optimization,” *IEEE Transactions on Automatic Control*, to appear, pp. 1–32, 2015.
- [23] B. Johansson, T. Keviczky, M. Johansson, and K. Johansson, “An interior-point Lagrangian decomposition method for separable convex optimization,” in *Proceedings of IEEE Conference on Decision and Control*, pp. 4185–4190, 2008.
- [24] I. Necoara and J. Suykens, “An interior-point Lagrangian decomposition method for separable convex optimization,” *Journal of Optimization Theory and Applications*, vol. 143, no. 3, pp. 567–588, 2009.
- [25] Y. Pu, M. Zeilinger, and C. Jones, “Quantization design for distributed optimization,” *Technical Report, EPF Lausanne*, pp. 1–15, 2015. [Online]. Available: http://infoscience.epfl.ch/record/207085/files/pu_quantization.pdf?version=1
- [26] E. Wei and A. Ozdaglar, “On the $1/k$ convergence if asynchronous alternating direction method of multipliers over networks,” *Technical Report, MIT*, pp. 1–30, 2015. [Online]. Available: <https://asu.mit.edu/sites/default/files/documents/publications/asynchronousADMM.pdf>
- [27] L. Carlone, V. Srivastava, F. Bullo, and G. Calafiore, “Distributed random convex programming via constraints consensus,” *SIAM Journal on Control and Optimization*, vol. 52, no. 1, pp. 629–662, 2014.
- [28] M. Bürger, G. Notarstefano, and F. Allgöwer, “A polyhedral approximation framework for convex and robust distributed optimization,” *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 3740–3754, 2014.
- [29] A. Nedic, A. Ozdaglar, and P. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
- [30] M. Zhu and S. Martinez, “On distributed convex optimization under inequality and equality constraints,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.
- [31] S. Lee and A. Nedic, “Distributed random projection algorithm for convex optimization,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 221–229, 2013.
- [32] K. Margellos, A. Falsone, S. Garatti, and M. Prandini, “Distributed constrained optimization and consensus in uncertain networks via proximal minimization,” *IEEE Transactions on Automatic Control*, under review, pp. 1–15, 2016. [Online]. Available: <http://arxiv.org/abs/1603.02239>
- [33] R. Bartle and D. Sherbert, *Introduction to real analysis*. John Wiley & Sons, Inc., fourth edition, 2011.
- [34] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, “On distributed averaging algorithms and quantization effects,” *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, 2009.
- [35] D. Bertsekas, “Multiplier methods: A survey,” *Automatica*, vol. 12, no. 2, pp. 133–145, 1976.
- [36] A. Nedic, A. Ozdaglar, and P. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *Technical Report*, pp. 1–35, 2008. [Online]. Available: <http://arxiv.org/abs/0802.3922v2>
- [37] J. Tsitsiklis and M. Athans, “Convergence and asymptotic agreement in distributed decision problems,” *IEEE Transactions on Automatic Control*, vol. 29, no. 1, pp. 42–50, 1984.
- [38] K. Margellos and S. Oren, “Capacity Controlled Demand Side Management: A Stochastic Pricing Analysis,” *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 706–717, 2016.