Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Subgradient Methods and Consensus Algorithms for Solving Convex Optimization Problems

Björn Johansson, Tamás Keviczky, Mikael Johansson, Karl Henrik Johansson

Abstract— **In this paper we propose a subgradient method for solving coupled optimization problems in a distributed way given restrictions on the communication topology. The iterative procedure maintains local variables at each node and relies on local subgradient updates in combination with a consensus process. The local subgradient steps are applied simultaneously as opposed to the standard sequential or cyclic procedure. We study convergence properties of the proposed scheme using results from consensus theory and approximate subgradient methods. The framework is illustrated on an optimal distributed finite-time rendezvous problem.**

I. INTRODUCTION

Large-scale networked systems are becoming ubiquitous and increasingly complex to manage. If the desired behavior of the networked system can be formulated as an optimization problem, then complexity issues can be mitigated by solving the resulting optimization problem with distributed algorithms. We will develop such a distributed algorithm.

Several important problems in applications such as resource allocation in computer networks [1], [2], estimation in sensor networks [3], and the rendezvous problem in multi-agent systems [4], can be posed as coupled optimization problems. In these problems, each node or agent in the network is associated with a component of the objective function, which depends on a network-wide decision variable.

Coupled optimization problems can be solved using a variety of distributed algorithms. A classical way is to use dual relaxation, i.e., the constraints are enforced using a pricing scheme. Another method is to iteratively refine an estimate of the optimizer using incremental subgradient methods, where the estimate typically is passed around in the network following a logical ring. In this paper, we combine consensus negotiations with a subgradient method, which is similar in flavor to the approach proposed in [5], but with different properties and analysis. In our approach, the communication topology is explicitly respected, and the algorithm only requires communication between neighboring nodes.

In Section II, we present our problem formulation as well as notation and connections with the existing literature. Then, in Section III, we introduce the algorithm and supporting lemmas. We then continue with a convergence analysis of the algorithm in Section IV. Furthermore, the properties of the algorithm is explored using a numerical example in Section V. Finally, we conclude the paper with a discussion in Section VI.

B. Johansson, M. Johansson and K.H. Johansson are with the ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology (KTH), 100 44 Stockholm, Sweden, {bjorn.johansson,mikaelj,kallej}@ee.kth.se.
Their work was supported by VR, SSF, VINNOVA, and HYCON.
T. Keviczky is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD, Delft, The Netherlands, t.keviczky@tudelft.nl

II. PROBLEM FORMULATION

In this paper we consider the following optimization problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) = \sum_{i=1}^{N} f^i(x) \\ \text{subject to} \quad & x \in \mathcal{X}, \end{aligned} \tag{1}$$

where $f^i : \mathbb{R}^M \to \mathbb{R}$ are convex functions and $\mathcal{X}$ is a nonempty, closed, and convex subset of $\mathbb{R}^M$. Let $f^\star$ denote the optimal value of (1) and let $x^\star$ denote an optimizer of (1). We will assume that in general f is nondifferentiable. Our interest in studying this class of optimization problems is motivated by optimal control problems for finite-time rendezvous of multiple dynamical agents [4], [6], resource allocation in computer networks [1], [2], and estimation in sensor networks [3].

A. Preliminaries

We will make use of the following definitions and assumptions in the paper.

Definition 1: A vector $g \in \mathbb{R}^M$ is a subgradient of a convex function $f : \mathbb{R}^M \to \mathbb{R}$ at a point $x \in \mathbb{R}^M$ if

$$f(y) \geq f(x) + g^\intercal (y - x), \quad \forall y \in \mathbb{R}^M. \tag{2}$$

Definition 2: The set of all subgradients of a convex function f at $x \in \mathbb{R}^M$ is called the subdifferential of f at x, and is denoted by $\partial f(x)$:

$$\partial f(x) = \left\{ g \in \mathbb{R}^M | f(y) \geq f(x) + g^\intercal (y - x), \forall y \in \mathbb{R}^M \right\}. \tag{3}$$

Definition 3: The ϵ-subdifferential set of a convex function f at x is the collection of ϵ-subgradients:

$$\begin{aligned} &\partial_\epsilon f(x) = \\ &\left\{ g \in \mathbb{R}^M | f(y) \geq f(x) + g^\intercal (y - x) - \epsilon, \forall y \in \mathbb{R}^M \right\}, \end{aligned} \tag{4}$$

with $\epsilon \geq 0$.

We will use $\|\cdot\|$ to denote the 2-norm of a vector x, i.e., $\|x\| = \sqrt{x^\intercal x}$, and the induced 2-norm of a matrix W, i.e., $\|W\| = \sup_{x \neq 0} \frac{\|Wx\|}{\|x\|}$.

Assumption 1 (Subgradient Boundedness): There exists a scalar C for all $i = 1, \ldots, N$ such that

$$\left\| g^i(x) \right\| \leq C, \quad \forall g^i(x) \in \partial f^i(x), \forall x \in \mathcal{X}. \tag{5}$$

B. Subgradient Methods

A popular method for solving problems of the type (1) is the subgradient method, which consists of the following iterative procedure

$$x_{k+1} = \mathcal{P}_{\mathcal{X}} \left[x_k + \alpha_k \sum_{i=1}^{N} g^i(x_k) \right], \tag{6}$$

where $g^i(x_k)$ is a subgradient of f^i at x_k, α_k is a positive stepsize, and $\mathcal{P}_{\mathcal{X}}$ denotes projection on the set $\mathcal{X} \subset \mathbb{R}^M$. References [7] and [8] contain excellent introductions and overviews of the theory and use of subgradient methods, as well as their convergence rate properties.

The subgradient method can be implemented in an incremental fashion as proposed in [8]. This entails changing the variable x_k incrementally through N steps, in each iteration using only the subgradient corresponding to a single component function f^i. The advantage of this method from a computational aspect is that it can be performed in a distributed way by assigning each component function to a processing unit or "agent", which performs the local incremental update on the variable x. This means that x_k needs to be "passed around" between the agents, which perform a subgradient update using only a single subgradient corresponding to the agent's component function. This incremental subgradient scheme has advantages over the standard one in terms of convergence rate and distribution of computations. However, in its usual form the incremental updates are performed in a sequential manner, which assumes that the variable x_k passes through all agents either in a cyclic or randomized sequence. Implementation of such a communication scheme can sometimes be problematic in practice.

A more realistic distribution of the computations would allow a much wider class of information exchange topologies, which might even be time-varying or suffer from delays.

Notice that we can think of the computing agents mentioned above as each having a copy of the decision variable x, which they maintain locally and update based on the value obtained from the previous subiteration (the preceding agent in the update sequence) and the local subgradient of the component function evaluated at this value. Under appropriate assumptions and using a properly chosen diminishing stepsize, the subgradient iterations converge asymptotically to an optimizer $x^\star$ of the problem. This means that eventually all "local" versions of the decision variable converge to the same value. This resembles to some extent agreement or consensus problems in multi-agent systems, which has a long history and has received renewed interest in the recent literature [9].

C. Consensus algorithms

The consensus problem considers conditions under which using a certain message-passing protocol, the local variables of each agent will converge to the same value. However, this value does not generally represent an optimizer of a problem of interest, and is typically a function of the initial values held by the agents and the information exchange policy.

A wide variety of results exist related to the convergence of local variables to a common value using various information exchange procedures among multiple agents [9]–[11]. It is thus interesting to investigate whether the convergence properties of certain consensus algorithms can be combined with subgradient iterations in order to optimize problems of type (1) using a wider variety of communication topologies than what the standard incremental subgradient method allows.

For our investigations we will consider the following consensus algorithm in discrete-time:

$$y_{k+1} = W y_k, \tag{7}$$

where the i-th element of vector y_k, denoted by $[y_k]_i$, corresponds to agent i. Furthermore, we associate an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with the optimization problem (1), where $\mathcal{V} = \{1, ..., N\}$ and $\mathcal{E} \subseteq V \times V$. The interpretation is that if $(i, j) \in \mathcal{E}$, then agent i can communicate with agent j. The information exchange among agents is represented by the matrix W, for which we make the following assumptions.

Assumption 2 (Consensus Matrix Properties): The weight matrix $W \in \mathbb{R}^{N \times N}$ fulfills

$$[W]_{ij} = 0, \text{ if } (i,j) \notin \mathcal{E} \text{ and } i \neq j,$$
$$W = W^\intercal,\ W\mathbf{1}_N = \mathbf{1}_N,\ \rho\left(W - \frac{\mathbf{1}_N \mathbf{1}_N^\intercal}{N}\right) \leq \gamma < 1,$$

where $\rho(\cdot)$ is the spectral radius and $\mathbf{1}_N \in \mathbb{R}^N$ is the column vector with all elements equal to one.

The matrix W can for example be chosen as the so-called Perron matrix of the communication graph $\mathcal{G}$ with parameter ε. It is defined as $W = I - \varepsilon L(\mathcal{G})$, where $L(\mathcal{G})$ represents the Laplacian matrix of the communication graph. If $\mathcal{G}$ with maximum degree Δ is strongly connected with $0 < \varepsilon < 1/\Delta$, then the limit $\lim_{k\to\infty} W^k$ exists and a consensus is asymptotically reached for all initial states. Furthermore, the consensus value will be the average of the initial states: $[y_\infty]_j = \frac{1}{N}\sum_{i=1}^{N}[y_0]_i$, for all $j = 1, ..., N$. For more details, see, e.g, [9]; for other ways of choosing W that fulfills Assumption 2, see, e.g, [11].

III. SUBGRADIENT ALGORITHM WITH CONSENSUS ITERATIONS

We propose to combine the subgradient iterations of (6) with a number of consensus iterations (7) in the following way:

$$x_{k+1}^i = \mathcal{P}_{\mathcal{X}} \left[\sum_{j=1}^{N} [W^\varphi]_{ij} \left(x_k^j - \alpha_k g^j(x_k^j) \right) \right], \tag{8}$$

where $g^j(x_k^j) \in \partial f^j(x_k^j)$ and $[W^\varphi]_{ij}$ denotes the element of W^φ in the i-th row and j-th column. Agents maintain their local variable x_k^i and perform the update procedure of (8) in parallel. First, they exchange information with neighboring agents for φ number of consensus iterations. More specifically, (8) implies that each agent i runs φ number of consensus iterations with its neighbors, defined by (7), for each row in the local vector x_k^i. Then, all agents implement the component subgradient update locally. The

total number of iterations in the algorithm up to step k is therefore $k\varphi$. For a more compact notation we define

$$u_k^i = x_k^i - \alpha_k g^i(x_k^i) \tag{9}$$

and

$$v_k^i = \sum_{j=1}^{N} [W^\varphi]_{ij} u_k^j,\ i = 1, ..., N. \tag{10}$$

Let us assume for the moment that in each subgradient iteration $\varphi \to \infty$, and the consensus updates converge to the average of the initial values (this holds if W fufills Assumption 2). Then, for all $i = 1, \ldots, N$ and $x_0^i \in \mathbb{R}^M$,

$$\lim_{\varphi\to\infty} \sum_{j=1}^{N} \left([W^\varphi]_{ij} u_0^j\right) = \frac{1}{N}\sum_{j=1}^{N}\left(x_0^j - \alpha_0 g^j(x_0^j)\right).$$

Let us denote the initial state of the projected consensus variable with

$$\breve{x}_1 = \mathcal{P}_{\mathcal{X}}\left[\frac{1}{N}\sum_{j=1}^{N}\left(x_0^j - \alpha_0 g^j(x_0^j)\right)\right].$$

In the next iteration, each agent will possess the same value $\breve{x}_1$, thus the local subgradients g^j will be evaluated at the same point. The update procedure in (8) for $k \geq 1$ will thus be equivalent to

$$\begin{aligned}\breve{x}_{k+1} &= \mathcal{P}_{\mathcal{X}}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\breve{x}_k - \alpha_k g^j(\breve{x}_k)\right)\right] \\ &= \mathcal{P}_{\mathcal{X}}\left[\breve{x}_k - \alpha_k \frac{1}{N}\sum_{j=1}^{N}(g^j(\breve{x}_k))\right].\end{aligned} \tag{11}$$

This is the same process as the standard subgradient method of (6) performed on the consensus variable $\breve{x}_k$ with a stepsize of α_k/N. Convergence analysis of this scheme can be done following the procedure in for example [7] or [8].

We are interested however in a more realistic scenario, where the consensus iterations are performed only for a finite number of φ steps. We intend to analyze such a scheme with the use of properly chosen approximate subgradients and the approximation error of the average consensus process achieved in a finite number of steps. Finally, we assume that the stepsize in (8) is constant, $\alpha_k = \alpha$, which makes the algorithm easily implementable.

The following three lemmas will be instrumental in the convergence analysis of the proposed scheme in Section IV. We will denote the average value of the local variables at time k with $\bar{x}_k = \frac{1}{N}\sum_{i=1}^{N} x_k^i$ and $\bar{v}_k = \frac{1}{N}\sum_{i=1}^{N} v_k^i$.

The following inequalities serve to characterize the Euclidean distance of agent variables from each other and from their average.

Lemma 1: 1) If $\left\|x_k^i - x_k^j\right\| \leq \beta$ for all $i, j = 1, ..., N$, then $\left\|x_k^j - \bar{x}_k\right\| = \left\|x_k^j - \frac{1}{N}\sum_{i=1}^{N} x_k^i\right\| \leq \frac{N-1}{N}\beta$.
2) If $\left\|x_k^i - \bar{x}_k\right\| \leq \beta$ for all $i = 1, ..., N$, then $\left\|x_k^i - x_k^j\right\| \leq 2\beta$.

Proof: 1) $\left\|x_k^j - \bar{x}_k\right\| = \frac{1}{N}\left\|N x_k^j - \sum_{i=1}^{N} x_k^i\right\| \leq \frac{1}{N}\left(\sum_{i=1, i\neq j}^{N}\left\|x_k^i - x_k^j\right\|\right) \leq \frac{N-1}{N}\beta$.
2) $\left\|x_k^i - x_k^j\right\| \leq \left\|x_k^i - \bar{x}\right\| + \left\|x_k^j - \bar{x}\right\| \leq 2\beta$. ■

Lemma 2: If $y_{k+1} = W^\varphi y_k$, with W fulfilling Assumption 2, and $\|[y_k]_i - [y_k]_j\| \leq \sigma$ for all $i, j = 1, ..., N$, then $\|[y_{k+1}]_i - [y_{k+1}]_j\| \leq 2\gamma^\varphi N\sigma$ for all $i, j = 1, ..., N$.

Proof: Let us write y_k as $y_k = \bar{y}_k + a_k$ with $\bar{y}_k = \mathbf{1}_N\mathbf{1}_N^\intercal y_k/N$ and $\sum_{i=1}^{N}[a_k]_i = 0$. The results of Lemma 1 show that $\|[a_k]_i\| \leq \sigma$ for all i. Furthermore, $y_{k+1} = W^\varphi(\bar{y}_k + a_k) = \bar{y}_k + W^\varphi(a_k - \mathbf{0}_N)$. Now we have,

$$\begin{aligned}&\|[y_{k+1} - \bar{y}_k]_i\| \leq \|y_{k+1} - \bar{y}_k\| = \|W^\varphi(a_k - \mathbf{0}_N)\| \\ &= \left\|\left(W^\varphi - \frac{\mathbf{1}_N\mathbf{1}_N^\intercal}{N}\right)a_k\right\| \leq \left\|W^\varphi - \left(\frac{\mathbf{1}_N\mathbf{1}_N^\intercal}{N}\right)^\varphi\right\|\|a_k\| \\ &\leq \left\|W - \frac{\mathbf{1}_N\mathbf{1}_N^\intercal}{N}\right\|^\varphi \|a_k\| \leq \gamma^\varphi\|a_k\| \leq \gamma^\varphi N\sigma,\end{aligned}$$

where we used that $\left\|W - \frac{\mathbf{1}_N\mathbf{1}_N^\intercal}{N}\right\| = \rho\left(W - \frac{\mathbf{1}_N\mathbf{1}_N^\intercal}{N}\right)$, which holds for symmetric matrices. Finally, from the above discussion and Lemma 1, we have $\|[y_{k+1}]_i - [y_{k+1}]_j\| \leq 2\gamma^\varphi N\sigma$ for all $i, j = 1, ..., N$. ■

The following lemma establishes a lower bound on the number of consensus steps that will ensure that the local variables will remain in a ball of radius β of their average, from one iteration to the next.

Lemma 3: Let $\{x_k^1, ..., x_k^N\}_{k=0}^{\infty}$ be generated by (8) under Assumptions 1 and 2. If $\left\|v_k^i - \bar{v}_k\right\| \leq \beta$ for all i and $\varphi \geq \left(\log(\beta) - \log(4MN(\beta + \alpha C))\right)/\log(\gamma)$, then $\left\|v_{k+1}^i - \bar{v}_{k+1}\right\| \leq \beta$ for all i.

Proof: Using Lemma 1, the closeness condition means that $\left\|v_k^i - v_k^j\right\| \leq 2\beta$ for all $i, j = 1, ..., N$, which implies that $\left\|\left[v_k^i - v_k^j\right]_l\right\| \leq 2\beta$ for all $i, j = 1, ..., N$ and $l = 1, ..., M$.

We can now decide the distance between the iterates before the consensus step.

$$\left\|\left[u_{k+1}^i\right]_l - \left[u_{k+1}^j\right]_l\right\| \tag{12a}$$

$$= \left\|\left[\mathcal{P}_{\mathcal{X}}[v_k^i] - \alpha g^i(\mathcal{P}_{\mathcal{X}}[v_k^i]) - \mathcal{P}_{\mathcal{X}}[v_k^j] + \alpha g^j(\mathcal{P}_{\mathcal{X}}[v_k^j])\right]_l\right\|$$

$$\leq \left\|\left[v_k^i - v_k^j\right]_l\right\| + 2\alpha C \leq 2(\beta + \alpha C), \tag{12b}$$

where we used the non-expansive property of the projection on a convex set. Furthermore, we have

$$\begin{aligned}&\left\|v_{k+1}^i - v_{k+1}^j\right\| \\ &= \left\|\sum_{n=1}^{N}[W^\varphi]_{in}u_{k+1}^n - \sum_{n=1}^{N}[W^\varphi]_{jn}u_{k+1}^n\right\| \\ &\leq \sum_{l=1}^{M}\left\|\left[\sum_{n=1}^{N}[W^\varphi]_{in}u_{k+1}^n - \sum_{n=1}^{N}[W^\varphi]_{jn}u_{k+1}^n\right]_l\right\|.\end{aligned}$$

Consider now one of the terms in the sum above and notice

that

$$\left\| \left[\sum_{n=1}^{N} [W^{\varphi}]_{in} u_{k+1}^{n} - \sum_{n=1}^{N} [W^{\varphi}]_{jn} u_{k+1}^{n} \right]_l \right\|$$
$$= \left\| \sum_{n=1}^{N} [W^{\varphi}]_{in} \left[u_{k+1}^{n}\right]_l - \sum_{n=1}^{N} [W^{\varphi}]_{jn} \left[u_{k+1}^{n}\right]_l \right\|$$
$$= \left\| [W^{\varphi} y_k]_i - [W^{\varphi} y_k]_j \right\|,$$

with $y_k = \left(\left[u_{k+1}^{1}\right]_l \quad \dots \quad \left[u_{k+1}^{N}\right]_l\right)^{\intercal}$. Using Lemma 2 and (12), which states that $\left\|[y_k]_i - [y_k]_j\right\| \leq 2(\beta + \alpha C)$, we obtain $\left\|[W^{\varphi} y_k]_i - [W^{\varphi} y_k]_j\right\| \leq 4\gamma^{\varphi} N(\beta + \alpha C)$.

Combining the above results yields $\left\|v_{k+1}^{i} - v_{k+1}^{j}\right\| \leq 4\gamma^{\varphi} MN(\beta + \alpha C)$, and $\left\|v_{k+1}^{i} - v_{k+1}^{j}\right\| \leq \beta$ is fulfilled if φ is set to

$$\varphi \geq \frac{\log(\beta) - \log(4MN(\beta + \alpha C))}{\log(\gamma)}.$$

■

IV. CONVERGENCE ANALYSIS

We will prove convergence of the algorithm in two cases: Firstly, convergence is proved when the feasible set is the space $\mathbb{R}^M$, i.e., the unconstrained optimization problem. Secondly, with an additional assumption on the objective functions, convergence is proved for a general convex feasible set.

A. Unconstrained Case

In the following, we will analyze the unconstrained case and we make the following assumption.

Assumption 3: The feasible set of (1) is $\mathcal{X} = \mathbb{R}^M$.

We need the following Lemma which allows us to interpret the algorithm as an ϵ-subgradient algorithm.

Lemma 4: Under Assumptions 1 and 2 and if $\left\|x_k^i - \bar{x}_k\right\| \leq \beta$ for all $i = 1, ..., N$, then $g^i(x_k^i) \in \partial_{\epsilon} f^i(\bar{x}_k)$ and $\sum_{i=1}^{N} g^i(x_k^i) \in \partial_{N\epsilon} f(\bar{x}_k)$, with $\epsilon = 2\beta C$.

Proof: Using the definition (2) and the bound (5) on the subgradient leads to

$$\begin{aligned} f^i(x_k^i) &\geq f^i(\bar{x}_k) + g^i(\bar{x}_k)^{\intercal}(x_k^i - \bar{x}_k) \\ &\geq f^i(\bar{x}_k) - \left\|g^i(\bar{x}_k)\right\| \left\|x_k^i - \bar{x}_k\right\| \\ &\geq f^i(\bar{x}_k) - C\beta. \end{aligned}$$

For any $y \in \mathcal{X}$, using the subgradient inequality leads to

$$\begin{aligned} f^i(y) &\geq f^i(x_k^i) + g^i(x_k^i)^{\intercal}(y - x_k^i) \\ &\geq f^i(\bar{x}_k) + g^i(x_k^i)^{\intercal}(y - x_k^i) - C\beta \\ &\geq f^i(\bar{x}_k) + g^i(x_k^i)^{\intercal}(y - \bar{x}_k + \bar{x}_k - x_k^i) - C\beta \\ &\geq f^i(\bar{x}_k) + g^i(x_k^i)^{\intercal}(y - \bar{x}_k) - 2C\beta. \end{aligned}$$

Using the definition of an ϵ-subdifferential (4), this implies $g^i(x_k^i) \in \partial_{2\beta C} f^i(\bar{x}_k)$. Summation of terms yields

$$f(y) \geq f(\bar{x}_k) + \left(\sum_{i=1}^{N} g^i(x_k^i)\right)^{\intercal}(y - \bar{x}_k) - N2\beta C.$$

Based on Definition 3, this implies $\sum_{i=1}^{N} g^i(x_k^i) \in \partial_{N2\beta C} f(\bar{x}_k)$. ■

Now we are ready for the convergence theorem for the unconstrained case.

Theorem 1: Under Assumptions 1, 2, and 3, with the sequence $\{x_k^1, ..., x_k^N\}_{k=0}^{\infty}$ generated by (8) with $\varphi \geq \big(\log(\beta) - \log(4NM(\beta + \alpha C))\big) / \log(\gamma)$ and $\left\|x_0^i - \bar{x}_0\right\| \leq \beta$, we have:
If $f^{\star} = -\infty$, then

$$\liminf_{k \to \infty} f(x_k^i) = -\infty, \forall i = 1, ..., N.$$

If $f^{\star} > -\infty$, then

$$\liminf_{k \to \infty} f(x_k^i) \leq f^{\star} + \alpha NC^2/2 + 3NC\beta, \forall i = 1, ..., N.$$

Proof: From Lemma 3, we know that $\left\|x_k^i - \bar{x}_k\right\| \leq \beta$ for all $i = 1, .., N$ and all $k \geq 0$, since $x_k^i = v_k^i$. Furthermore, from Lemma 4, we know that $\sum_{i=1}^{N} g^i(x_k^i) \in \partial_{2N\beta C} f(\bar{x}_k)$ for all $k \geq 0$. Hence, from the definitions of $\bar{x}_k$ and x_k^i in combination with the results above, we have

$$\begin{aligned} \bar{x}_{k+1} &= \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} [W^{\varphi}]_{ij} \left(x_k^j - \alpha g^j(x_k^j) \right) \right) = \\ &= \frac{1}{N} \sum_{j=1}^{N} \left(x_k^j - \alpha g^j(x_k^j) \right) = \bar{x}_k + \frac{\alpha}{N} h(\bar{x}_k), \end{aligned}$$

with $h(\bar{x}_k) \in \partial_{2N\beta C} f(\bar{x}_k)$ and $\|h(\bar{x}_k)\| \leq NC$. This is precisely the approximate subgradient iteration and from [12, Proposition 4.1] we have

$$\liminf_{k \to \infty} f(\bar{x}_k) = -\infty, \text{ if } f^{\star} = -\infty$$

and

$$\liminf_{k \to \infty} f(\bar{x}_k) \leq f^{\star} + \alpha(NC)^2/(2N) + 2NC\beta, \text{ if } f^{\star} > -\infty.$$

By noting that

$$f(x_k^i) \leq f(\bar{x}_k) + NC\beta, \forall i = 1, ..., N, k \geq 0,$$

we have the desired result. ■

Remark 1: We can get $\liminf_{k \to \infty} f(x_k^i)$ to be arbitrarily close to $f^{\star}$, by choosing the constants α and β arbitrarily small. Note that the number of required consensus negotiations (φ) to reach a fixed β does not depend on k.

B. Constrained Case

To show convergence in the constrained case, we need the following additional assumption on the functions f^i.

Assumption 4: There exist $\zeta > 0$ and $\tau > 0$ such that for all $x \in \mathcal{X}$, $g^i(x) \in \partial f^i(x)$, and $\nu \in \mathbb{R}^M$ with $\|\nu\| \leq \tau$, the following holds:

$$g^i(x) + \nu \in \partial_{\zeta} f^i(x).$$

As mentioned before, this is an additional assumption compared to what is needed for the unconstrained case. However, if $\mathcal{X}$ is a compact set, e.g., $x \in \mathcal{X} \Rightarrow \|x\| \leq \eta$, then Assumption 4 is fulfilled for convex functions and arbitrarily

small (but fixed) ζ as long as τ is sufficiently small. To see this, consider $x, w \in \mathcal{X}$, then

$$\begin{aligned} f^i(w) &\geq f^i(x) + g^i(x)^\intercal(w-x) \\ &= f^i(x) + (g^i(x) + \nu - \nu)^\intercal(w-x) \\ &\geq f^i(x) + (g^i(x)+\nu)^\intercal(w-x) - \|\nu\|\left(\|x\| + \|w\|\right) \\ &\geq f^i(x) + (g^i(x)+\nu)^\intercal(w-x) - 2\tau\eta, \end{aligned}$$

which fulfills Assumption 4 if $\tau \leq \zeta/(2\eta)$. Another example is when $f(x) = \psi x^\intercal x$, then Assumption 4 is fulfilled if $\tau \leq 2\sqrt{\zeta\psi}$ (without $\mathcal{X}$ necessarily being a compact set).

In our further developments we need to keep track of the difference $\bar{x}_{k+1} - \mathcal{P}_{\mathcal{X}}\left[\bar{v}_k\right]$, and to this end, we define y_k and z_k as

$$y_k = \mathcal{P}_{\mathcal{X}}\left[\bar{v}_{k-1}\right] \text{ and } z_k = \bar{x}_k - y_k. \tag{17}$$

Furthermore, we need the following Lemma, which is similar to Lemma 4.

Lemma 5: Under Assumptions 1, 2, and 4 and if $\left\|v^i_{k-1} - \bar{v}_{k-1}\right\| \leq \beta$ for all i and $\nu \in \mathbb{R}^M$ with $\|\nu\| \leq \tau$, then $(g^i(x^i_k) + \nu) \in \partial_\epsilon f^i(y_k)$ and $\sum_{i=1}^N (g^i(x^i_k) + \nu) \in \partial_{N\epsilon} f^i(y_k)$ with $\epsilon = \beta(6C + 3\tau) + \zeta$.

Proof: The proof idea relies on the iterates being close to each other. Using the assumptions and the non-expansive property of projection on a convex set we can bound the distance $\|z_k\|$ as follows

$$\begin{aligned} \|z_k\| &= \frac{1}{N}\left\|\sum_{i=1}^N \left(\mathcal{P}_{\mathcal{X}}\left[v^i_{k-1}\right] - \mathcal{P}_{\mathcal{X}}\left[\bar{v}_{k-1}\right]\right)\right\| \\ &\leq \frac{1}{N}\sum_{i=1}^N \left\|v^i_{k-1} - \bar{v}_{k-1}\right\| \leq \beta. \end{aligned}$$

Using the definition (2), the bound on the subgradient (5), the bound above, and Lemma 1 we obtain

$$\begin{aligned} f^i(x^i_k) &\geq f^i(y_k) + g^i(y_k)^\intercal(x^i_k - y_k) \\ &\geq f^i(y_k) - \left\|g^i(y_k)\right\|\left\|x^i_k - y_k\right\| \\ &\geq f^i(y_k) - C(\left\|x^i_k - \bar{x}_k\right\| + \|z_k\|) \\ &\geq f^i(y_k) - C3\beta, \end{aligned}$$

where we used $x^i_k - y_k = x^i_k - \bar{x}_k + z_k$ and $\left\|v^i_k - \bar{v}_k\right\| \leq \beta \Rightarrow \left\|v^i_k - v^j_k\right\| \leq 2\beta \Rightarrow \left\|x^i_k - x^j_k\right\| \leq 2\beta \Rightarrow \left\|x^i_k - \bar{x}_k\right\| \leq 2\beta$.

For any $y \in \mathcal{X}$, using Assumption 4 and the previous arguments, we get

$$\begin{aligned} &f^i(y) \geq f^i(x^i_k) + (g^i(x^i_k) + \nu)^\intercal(y - x^i_k) - \zeta \\ &\geq f^i(y_k) + (g^i(x^i_k) + \nu)^\intercal(y - x^i) - (3C\beta + \zeta) \\ &= f^i(y_k) + (g^i(x^i_k) + \nu)^\intercal(y - y_k + y_k - x^i_k) - (3C\beta + \zeta) \\ &\geq f^i(y_k) + (g^i(x^i_k) + \nu)^\intercal(y - y_k) - (\beta(6C + 3\tau) + \zeta). \end{aligned}$$

Using the definition of an ϵ-subdifferential (4), this implies $(g^i(x^i_k) + \nu) \in \partial_{(\beta(6C+3\tau)+\zeta)} f^i(y_k)$. Summation of terms yields

$$\begin{aligned} &f(y) \geq f(y_k) + \\ &\left(\sum_{i=1}^N (g^i(x^i_k) + \nu)\right)^\intercal (y - y_k) - N(\beta(6C + \tau) + \zeta). \end{aligned}$$

Based on Definition 3, this implies $\sum_{i=1}^N (g^i(x^i_k) + \nu) \in \partial_{N(\beta(6C+3\tau)+\zeta)} f(y_k)$. ■

We are now ready for the convergence theorem, which is based on the idea of interpreting (8) as an approximate subgradient algorithm.

Theorem 2: Under Assumptions 1; 2; and 4, with the sequence $\{x^1_k, ..., x^N_k\}_{k=0}^\infty$ generated by (8) with $\varphi \geq \left(\log(\beta) - \log(4NM(\beta + \alpha C))\right)/\log(\gamma)$; $\left\|v^i_0 - \bar{v}_0\right\| \leq \beta$; and $\beta/\alpha \leq \tau$, we have:
If $f^\star = -\infty$, then

$$\liminf_{k\to\infty} f(x^i_k) = -\infty, \forall i = 1, ..., N.$$

If $f^\star > -\infty$, then

$$\begin{aligned} \liminf_{k\to\infty} f(x^i_k) \leq f^\star + \alpha N(C+\tau)^2/2 + \\ N(\beta(9C + 3\tau) + \zeta), \forall i = 1, ..., N. \end{aligned}$$

Proof: From the definition of y_k we have,

$$\begin{aligned} y_{k+1} &= \mathcal{P}_{\mathcal{X}}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^N v^i_k\right] \\ &= \mathcal{P}_{\mathcal{X}}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^N \left(x^i_k - \alpha g^i(x^i_k)\right)\right] \\ &= \mathcal{P}_{\mathcal{X}}\left[y_k + z_k - \tfrac{\alpha}{N}\textstyle\sum_{i=1}^N g^i(x^i_k)\right] \\ &= \mathcal{P}_{\mathcal{X}}\left[y_k - \tfrac{\alpha}{N}\textstyle\sum_{i=1}^N \left(g^i(x^i_k) - \tfrac{z_k}{\alpha}\right)\right]. \end{aligned}$$

From Lemma 3, we know that $\left\|v^i_k - \bar{v}_k\right\| \leq \beta$ for all $i = 1, .., N$ and all $k \geq 0$. Furthermore, from Lemma 5, we know that $\sum_{i=1}^N (g^i(x^i_k) - z_k/\alpha) \in \partial_{N(\beta(6C+3\tau)+\zeta)} f(y_k)$ for all $k \geq 1$, since $\|z_k/\alpha\| \leq \beta/\alpha \leq \tau$ by assumption. In addition, we know that $\left\|\sum_{i=1}^N (g^i(x^i_k) - z_k/\alpha)\right\| \leq N(C+\tau)$. Hence, y_k is updated according to an approximate subgradient method and from [12, Proposition 4.1] we have

$$\liminf_{k\to\infty} f(y_k) = -\infty, \text{ if } f^\star = -\infty$$

and

$$\begin{aligned} \liminf_{k\to\infty} f(y_k) \leq f^\star + \left(\alpha(N(C+\tau))^2\right)/(2N) + \\ N(\beta(6C + 3\tau) + \zeta), \text{ if } f^\star > -\infty. \end{aligned}$$

Finally, we have

$$\begin{aligned} f(x^i_k) &\leq f(y_k) + N\beta 3C \\ &\leq f^\star + \alpha N(C+\tau)^2/2 + \\ &\quad N(\beta(9C + 3\tau) + \zeta), \forall i = 1, ..., N, k \geq 0. \end{aligned}$$

■

Remark 2: The assumption $\beta \leq \alpha\tau$ in Theorem 2 may seem restrictive, but it can be fulfilled with a fixed number of consensus negotiations according to Lemma 3.

Remark 3: The initial conditions in Theorem 1 and Theorem 2, $\left\|x^i_0 - \bar{x}_0\right\| \leq \beta$ and $\left\|v^i_0 - \bar{v}_0\right\| \leq \beta$, respectively, can be fulfilled with sufficiently many consensus negotiations before starting the algorithm. Another simple alternative is to set $x^i_0 = x^j_0, \forall i, j$ and $v^i_0 = v^j_0, \forall i, j$, respectively.

V. NUMERICAL EXAMPLE

The proposed subgradient algorithm in combination with consensus iterates as shown in (8) has been implemented to solve a distributed finite-time optimal rendezvous problem involving three agents moving in a plane with double integrator dynamics. A more detailed problem description can be found in [4]. Each agent maintains a local version of the optimization variable x_k^i (the rendezvous position) and performs the local subgradient update with a diminishing stepsize according to (8). The consensus updates are performed with the following matrix W:

$$W = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.25 & 0.75 \end{bmatrix}. \tag{20}$$

Figure 1 illustrates the convergence of the local optimization variable for one of the agents to the globally optimal rendezvous location. The evolution of the local variable for the other agents is very similar. The different curves overlaid in the graph correspond to different choices of the consensus iteration limit (i.e., the maximum number φ of consensus steps performed). This illustrates how increased communication in terms of the number of consensus iterations with neighboring agents leads to better convergence rate. As we increase φ we expect to approach the convergence rate of the standard subgradient method given in (6). The main advantage of our proposed scheme is that the subgradient updates are performed in parallel using information from neighboring agents and there is no need for a central processing unit to whom each agent should communicate its local subgradient.

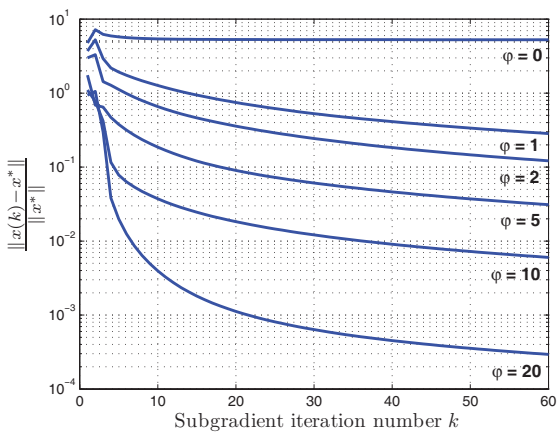

Fig. 1. Convergence plots for increasing number of consensus iterations (φ).

VI. CONCLUSIONS AND FUTURE WORK

In this paper we have described an iterative subgradient-based method for solving coupled optimization problems in a distributed way given restrictions on the communication topology. In order to allow great flexibility in the information exchange architecture and distribute calculations, we combined the local subgradient updates with a consensus process. This means that computing agents can work in parallel with each other and use localized information exchange. The generality of results in consensus theory promises relatively easy extensions of the presented method to situations where the interconnection topology might be time-varying, delayed and the local updates are performed asynchronously. Research on the speed of convergence in consensus protocols has an immediate and clear use in our framework.

For analysis purposes, we used results from consensus theory and employed approximate subgradient methods to study convergence properties of the proposed scheme. A connection is established between the number of consensus steps and the resulting level of optimality obtained by the subgradient updates. We have illustrated the effect of choosing different consensus iteration limits in an optimal distributed finite-time rendezvous problem.

A different version of (8) is also conceivable as

$$x_{k+1}^i = \mathcal{P}_{\mathcal{X}}\left[\left(\sum_{j=1}^N [W^\varphi]_{ij} x_k^j\right) - \alpha_k g^i\left(\sum_{j=1}^N [W^\varphi]_{ij} x_k^j\right)\right],$$

where in each update first a consensus is reached on the local variables (at least approximately), and then the corresponding local subgradient updates are calculated and applied. We expect that a similar line of analysis shown in this paper can be performed for such a variation of the original scenario.

We are currently exploring different choices of the stepsize, e.g., diminishing stepsizes, and we believe that other results from approximate subgradient methods can be used. Finally, we are also looking into how the different parameters should be tuned, i.e., should we decrease the stepsize or increase the number of negotiations to get closer to optimality?

REFERENCES

[1] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[2] B. Johansson, P. Soldati, and M. Johansson, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1535–1547, Aug. 2006.

[3] M. Rabbat, R. Nowak, and J. Bucklew, "Generalized consensus computation in networked systems with erasure links," in *IEEE SPAWC*, 2005.

[4] B. Johansson, A. Speranzon, M. Johansson, and K. H. Johansson, "On decentralized negotiation of optimal consensus," *Automatica*, pp. 1175–1179, Jul. 2008.

[5] A. Nedić and A. Ozdaglar, "On the rate of convergence of distributed asynchronous subgradient methods for multi-agent optimization," in *IEEE CDC*, 2007.

[6] T. Keviczky and K. H. Johansson, "A study on distributed model predictive consensus," in *IFAC World Congress*, 2008.

[7] K. C. Kiwiel, "Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM J. Optim.*, vol. 14, no. 3, pp. 807–840, 2004.

[8] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.

[9] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[10] A. Olshevsky and J. Tsitsiklis, "Convergence rates in distributed consensus and averaging," in *IEEE CDC*, 2006.

[11] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[12] A. Nedić, "Subgradient methods for convex minimization," Ph.D. dissertation, MIT, 2002.