2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

Distributed Nonlinear Programming Methods for Optimization Problems with Inequality Constraints

Ion Matei, John S. Baras

Abstract— **In this paper we consider a distributed optimization problem, where a set of agents interacting and cooperating locally have as common goal the minimization of a function expressed as a sum of (possibly non-convex) differentiable functions. Each function in the sum is associated with an agent and each agent has assigned an inequality constraint, therefore generating an optimization problem with inequality constraints. In this paper we present a distributed algorithm for solving such a problem, and give local convergence results. Our approach is based on solving (in a centralized manner) an equivalent augmented optimization problem with mixed constraints. The structure of this augmented problem ensures that the resulting algorithm is distributed. The main challenge in proving the convergence results comes from the fact that the local minimizers are no longer regular due to the distributed formulation. We present also an extension of this algorithm that solves a constrained optimization problem, where each agent has both equality and inequality constraints.**

I. Introduction

Multi-agent, distributed optimization algorithms solve problems where a group of agents has as common goal the optimization of a cost function under limited information and resources. The limited information is usually induced by agents being able to communicate with only a subset of the total set of agents. The authors of [17] introduced a multi-agent, distributed optimization algorithm, where the convex optimization cost is expressed as a sum of functions and each function in the sum corresponds to an agent. The agents cooperation is conditioned by a communication network, usually modeled as an undirected graph. The algorithm combines a (sub)gradient descent step with a consensus step; the latter being added to deal with the fact that the agents have only limited information about the cost function. Extensions of this initial version followed in the literature. [15], [18] include communication noise and errors on subgradients, [10], [12] assume a random communication graph, [15], [20] study asynchronous versions of the algorithm, [11] considers state-dependent communication topologies, while [3] assumes directed communication graphs. Another modification of the algorithm described in [17] was introduced in [8], where the authors change the order in which the consensus-step and the subgradient descent step are executed. Algorithms of the same flavor were also used to solve convex, constrained, optimization problems where the agents share a global constraint set [9], [15], [18] or where each agent has its own set of constraints [16], [20]. Other approaches for obtaining distributed algorithms use dual decomposition [21], augmented Lagragian [6], [7], or in particular, distributed versions of the Alternating Direction Method of Multipliers (ADMM) algorithm [2], [19], [22].

In this paper, we extend nonlinear programming techniques to a distributed optimization setup. In particular, the objective function expressed as a sum of functions, each function being associated with an agent. In addition, each agent has an inequality constraint assigned to it, as well. We propose a distributed algorithm derived from an algorithm used to solve an augmented constrained optimization problem in a centralized manner. Our approach is based on first formulating an equivalent augmented optimization problem with mixed constraints; a problem that is again reformulated so that it contains only equality constraints. Using a centralized first order algorithm to solve the first order necessary condition for the latter augmented optimization problem, we show that in fact we solve the original constrained optimization problem in a distributed manner. In our formulation we make no convexity assumptions on the cost and constraint functions, but we assume they are continuously differentiable. As a consequence our convergence results are local, and the main challenge in proving the results focuses on dealing with the effect of the non-regularity of the local minimizers, due to the distributed formulation. Distributed algorithms for solving constrained, non-convex optimization problems were proposed in [13], [14] and [23]. In particular, [23] solves also a non-convex optimization problem with inequality constraints. Note however that the inequality constraints are assumed globally known and their approach is based on solving an approximate version of the original problem.

The paper is organized as follows: in Section II we introduce the problem setup. Section III formulates an equivalent augmented optimization problem with mixed constraints and presents results concerning the equivalence between the two problems. In Section IV we re-formulate the augmented problem into a problem that contains only equality constraints. This section introduces also a set of results on properties of the Lagrangian of the two augmented problems; properties used for proving the convergence results. Section V presents the convergence results of a centralized algorithm for solving the augmented optimization problem with equality constraints, while Section VI shows how in fact this centralized algorithm is a distributed algorithm for

Ion Matei, is with the System Sciences Laboratory at Palo Alto Research Center (PARC), Palo Alto, CA (emails: ion.matei@parc.com); John S. Baras is with the Institute for Systems Research at the University of Maryland, College Park, MD (email: baras@umd.edu).

Research partially supported by US Air Force Office of Scientific Research MURI grant FA9550-09-1-0538, by National Science Foundation (NSF) grant CNS-1035655, and by National Institute of Standards and Technology (NIST) grant 70NANB11H148 to the University of Maryland.

obtaining the solution of the original problem. In addition, in this section we introduce an extension of the algorithm that is applied in the case agents have associated both equality and inequality constraints.

Notations and definitions: For a matrix A, its (i,j) entry is denoted by $[A]_{ij}$ and its transpose is given by A'. If A is a symmetric matrix, $A \succ 0$ ($A \succeq 0$) means that A is positive (semi-positive) definite. The nullspace and range of A are denoted by Null(A) and Range(A), respectively. The symbol $\otimes$ is used to represent the Kronecker product between two matrices. The vector of all ones is denoted by $\mathbb{1}$. Let $\{A_i\}_{i=1}^N$ be a set of matrices. By diag(A_i, $i = 1,\ldots,N$) we understand a block diagonal matrix, where the i^{th} block matrix is given by A_i. By $\mathcal{S}(x;\varepsilon)$ we refer to a neighborhood around x, of radius ε. Let $f:\mathbb{R}^n \to \mathbb{R}$ be a function. We denote by $\nabla f(x)$ and by $\nabla^2 f(x)$ the gradient and the Hessian of f at x, respectively. Let $F:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a function of variables (x,y). The block descriptions of the gradient and of the Hessian of F at (x,y) are given by $\nabla F(x,y)' = \left(\nabla_x F(x,y)', \nabla_y F(x,y)'\right)$, and

$$\nabla^2 F(x,y) = \begin{pmatrix} \nabla^2_{xx}F(x,y) & \nabla^2_{xy}F(x,y) \\ \nabla^2_{xy}F(x,y) & \nabla^2_{yy}F(x,y) \end{pmatrix},$$

respectively.

II. Problem description

In this section we describe the setup of our problem. We present first the communication model followed by the optimization model.

A. Communication model

A set of N agents interact with each other through a communication topology modeled as an undirected communication graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, where $\mathcal{V} = \{1,2,\ldots,N\}$ is the set of nodes and $\mathcal{E} = \{e_{ij}\}$ is the set of edges. An edge between two nodes i and j means that agents i and j can exchange information (or can cooperate). We assume that at each time instant k the agents can synchronously exchange information with their neighbors. We denote by $\mathcal{N}_i \triangleq \{j \mid e_{ij} \in \mathcal{E}\}$ the set of neighbors of agent i. We denote by $L \in \mathbb{R}^{N\times N}$ the (weighted) Laplacian of the graph $\mathcal{G}$, defined as

$$[L]_{ij} = \begin{cases} -l_{ij} & j \in \mathcal{N}_i, \\ \sum_{j\in\mathcal{N}_i} l_{ij} & j = i, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where l_{ij} are given positive scalars.

In the next sections we are going to make use of a set of properties of the matrix L; properties that are grouped in the following remark.

Proposition 2.1: The matrix L defined with respect to a connected graph $\mathcal{G}$ satisfies the following properties:

(a) The nullspace of L is given by Null(L) $= \{\gamma\mathbb{1} \mid \gamma \in \mathbb{R}\}$;
(b) Let $\mathbf{L} = L \otimes I$, where I is the n-dimensional identity matrix. Then the nullspace of $\mathbf{L}$ is given by Null($\mathbf{L}$) $= \{\mathbb{1} \otimes x \mid x \in \mathbb{R}^n\}$.□

B. Optimization model

We consider a function $f:\mathbb{R}^n \to \mathbb{R}$ expressed as a sum of N functions $f(x) = \sum_{i=1}^N f_i(x)$, and a vector-valued function $g:\mathbb{R}^n \to \mathbb{R}^N$ where $g \triangleq (g_1, g_2, \ldots, g_N)'$, with $g_i:\mathbb{R}^n \to \mathbb{R}$ and $N \leq n$.

We make the following assumptions on the functions f and g and on the communication model.

Assumption 2.1: (a) The functions $f_i(x)$ and $g_i(x)$, $i = 1,\ldots,N$ are twice continuously differentiable;
(b) Agent i has knowledge of only functions $f_i(x)$ and $g_i(x)$, and scalars l_{ij}, for $j \in \mathcal{N}_i$;
(c) Agent i can exchange information only with agents in the set of neighbors defined by $\mathcal{N}_i$;
(d) The communication graph $\mathcal{G}$ is connected.

The common goal of the agents is to solve the following optimization problem with equality constraints

$$(P_1) \quad \min_{x\in\mathbb{R}^n} \quad f(x), \qquad \text{subject to:} \quad g(x) \leq 0,$$

under Assumptions 2.1, where the inequality is entry-wise. Throughout the rest of the paper we assume that problem (P_1) has at least one local minimizer.

Let x^* be a local minimizer of (P_1) and let $B(x^*)$ be the set of indices corresponding to the active constraints, that is, $B(x^*) = \{i \mid g_i(x^*) = 0\}$, and let $\nabla g_i(x^*)$ denote the gradient of $g_i(x)$ at x^*. The following assumption is used to guarantee the uniqueness of the Lagrange multiplier vector ψ^* appearing in the Kuhn-Tucker necessary conditions of (P_1).

Assumption 2.2: Let x^* be a local minimizer of (P_1). The vectors $\{\nabla g_i(x^*)\}_{i\in B(x^*)}$ are linearly independent.

Under such assumption the Lagrange multiplier vector ψ^* such that $\nabla f(x^*) + \sum_{i=1}^N \psi_i^* \nabla g_i(x^*) = 0$, with $\psi_i^* \geq 0$ for all i, and $\psi_i^* = 0$ for all $i \notin B(x^*)$, is unique (see for example Section 3.3, page 283 of [1])

Assumption 2.2 is typically used to prove local convergence for several algorithms for solving (P_1). As we will see in the next sections, the same assumption will be used to prove local convergence for a distributed algorithm used to solve an augmented optimization problem with inequality constraints.

As seen later in the paper, the approach chosen for dealing with the inequality constraints is to transform them into equality constraints by introducing additional variables:

$$(P_1^*) \quad \min_{x,z} \quad f(x), \qquad \text{subject to:} \quad g_i(x) + z_i^2 = 0,\ i = 1\ldots,N,$$

where $z' = (z_1,\ldots,z_N)$. It can be shown (Section 3.3.2, page 286, [1]) that under Assumptions 2.1-(a) and 2.2, indeed solving (P_1^*) is equivalent to solving (P_1).

III. An equivalent optimization problem with mixed constraints

In this section we define an augmented optimization problem, from whose solution we can in fact extract the solution of problem (P_1). This approach will allow us to use centralized algorithms which result in distributed algorithms for (P_1) by leveraging the structure of the equivalent augmented problem.

Let us define the function $\mathbf{F}:\mathbb{R}^{nN}\to\mathbb{R}$ given by $\mathbf{F}(\mathbf{x})=\sum_{i=1}^N f_i(x_i)$, where $\mathbf{x}'=(x_1',x_2',\ldots,x_N')$, with $x_i\in\mathbb{R}^n$. In addition we introduce the vector-valued functions $\mathbf{g}:\mathbb{R}^{nN}\to\mathbb{R}^N$ and $\mathbf{h}:\mathbb{R}^{nN}\to\mathbb{R}^{nN}$, where $\mathbf{g}(\mathbf{x})=(\mathbf{g}_1(\mathbf{x}),\mathbf{g}_2(\mathbf{x}),\ldots,\mathbf{g}_N(\mathbf{x}))'$, with $\mathbf{g}_i:\mathbb{R}^{nN}\to\mathbb{R}$ given by $\mathbf{g}_i(\mathbf{x})=g_i(x_i)$, and $\mathbf{h}(\mathbf{x})'=(h_1(\mathbf{x})',h_2(\mathbf{x})',\ldots,h_N(\mathbf{x})')$, with $h_i:\mathbb{R}^{n\bar{N}}\to\mathbb{R}^n$ given by $h_i(\mathbf{x})=\sum_{j\in\mathcal{N}_i}l_{ij}(x_i-x_j)$, with l_{ij} positive scalars. The vector valued function $\mathbf{h}(\mathbf{x})$ can be compactly expressed as $\mathbf{h}(\mathbf{x})=\mathbf{L}\mathbf{x}$, where $\mathbf{L}=L\otimes I$, with I the n-dimensional identity matrix and L defined in (1). We introduce the optimization problem

$$(P_2)\qquad \min_{\mathbf{x}\in\mathbb{R}^{nN}}\quad \mathbf{F}(\mathbf{x}), \tag{2}$$

$$\text{subject to:}\quad \mathbf{h}(\mathbf{x})=0, \tag{3}$$

$$\mathbf{g}(\mathbf{x})\leq 0. \tag{4}$$

The Lagrangian function of problem (P_2) is a function $\mathcal{L}:\mathbb{R}^{nN}\times\mathbb{R}^N\times\mathbb{R}^{n\bar{N}}\to\mathbb{R}$, defined as

$$\mathcal{L}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda})\triangleq\mathbf{F}(\mathbf{x})+\boldsymbol{\mu}'\mathbf{g}(\mathbf{x})+\boldsymbol{\lambda}'\mathbf{L}\mathbf{x}. \tag{5}$$

The vectors $\nabla\mathbf{g}_i(\mathbf{x})$ are the gradients of the functions $\mathbf{g}_i(\mathbf{x})$ with a structure given by

$$\nabla\mathbf{g}_i(\mathbf{x}^*)'=\left[\underbrace{0,\ldots,0}_{n\text{ zeros}},\ldots,\underbrace{\nabla g_i(x^*)'}_{i^{\text{th}}\text{ component}},\ldots,\underbrace{0,\ldots,0}_{n\text{ zeros}}\right], \tag{6}$$

as per definition of the function $\mathbf{g}_i(\mathbf{x})$.

Remark 3.1: If $\mathbf{x}^*$ is a local minimizer of (P_1), then it follows immediately that if $\{\nabla g_i(x^*)\}_{i\in B(x^*)}$ is a set of linearly independent vectors then $\{\nabla\mathbf{g}_i(\mathbf{x}^*)\}_{i\in\boldsymbol{B}(\mathbf{x}^*)}$ is also a set of linearly independent vectors, where $\mathbf{x}^*=\mathbb{1}\otimes x^*$ and $B(\mathbf{x}^*)$ is the set of active constraints of (P_2); that is, $\boldsymbol{B}(\mathbf{x}^*)=\{i\mid\mathbf{g}_i(\mathbf{x}^*)=0\}$. In addition, it can be easily seen that $B(x^*)=\boldsymbol{B}(\mathbf{x}^*)$.

The following proposition states the necessary conditions that a local minimizer-Lagrange multipliers pair of (P_2) must satisfy.

Proposition 3.1 (first order necessary conditions): Let Assumptions 2.1 and 2.2 hold and let $\mathbf{x}^*=\mathbb{1}\otimes x^*$ be a local minimizer for problem (P_2) satisfying the constraints $\mathbf{g}(\mathbf{x}^*)\leq 0$ and $\mathbf{h}(\mathbf{x}^*)=0$. There exist unique vectors $\boldsymbol{\mu}^*$ and $\boldsymbol{\lambda}^*\in\text{Range}(\mathbf{L})$ so that $\nabla\mathbf{F}(\mathbf{x}^*)+\sum_{i=1}^N\boldsymbol{\mu}_i^*\nabla\mathbf{g}_i(\mathbf{x}^*)+\mathbf{L}'\tilde{\boldsymbol{\lambda}}=0$ for all $\tilde{\boldsymbol{\lambda}}\in\{\boldsymbol{\lambda}^*+\boldsymbol{\lambda}_\perp\mid\boldsymbol{\lambda}_\perp\in\text{Null}(\mathbf{L}')\}$. In addition, $\boldsymbol{\mu}_i^*\geq 0$ for all i, $\boldsymbol{\mu}_i^*=0$ for all $i\notin\boldsymbol{B}(\mathbf{x}^*)$ and $\mathbf{y}'\nabla_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\mathbf{y}\geq 0$ for all $\mathbf{y}$ such that $\mathbf{L}\mathbf{y}=0$ and $\nabla\mathbf{g}_i(\mathbf{x}^*)'\mathbf{y}=0$ for all $i\in\boldsymbol{B}(\mathbf{x}^*)$.□

Note that due to the fact that L (and consequently $\mathbf{L}$) is not full rank, the Lagrange multiplier $\boldsymbol{\lambda}$ verifying the necessary conditions is no longer unique. Still its particular structure will be exploited in what follows.

The following proposition states that by solving (P_2) we solve in fact (P_1) as well, and vice-versa.

Proposition 3.2: Let Assumptions 2.1 hold. The vector x^* is a local minimizer of (P_1) if and only if $\mathbf{x}^*=\mathbb{1}\otimes x^*$ is a local minimizer of (P_2). □

Remark 3.2: Note that any feasible solution of (P_2) must satisfy constraint (3) and therefore the assumption of a connected topology makes sure that any solution is of the form $\mathbf{x}^*=\mathbb{1}\otimes x^*$. The fact that we search a solution of (P_2) with this particular structure is fundamental for showing the equivalence between the two optimization problems. □

Under the assumption that $\{\nabla g_i(x^*)\}_{i\in B(x^*)}$ are linearly independent, it is well known (see for example Proposition 3.3.1, page 284, [1]) that there exists a unique vector $\psi^*=(\psi_i^*)$ satisfying the necessary conditions of (P_1), namely $\nabla f(x^*)+\sum_i\psi_i^*\nabla g_i(x^*)=0$. Without much effort it can be shown that this vector can be directly recovered from the solution of $(P2)$, as stated in the following proposition.

Proposition 3.3: Let Assumptions 2.1 and 2.2 hold, let $\mathbf{x}^*=\mathbb{1}\otimes x^*$ be a local minimizer of (P_2) and let $\psi^*=(\psi_i^*)$ and $\boldsymbol{\mu}^*=(\boldsymbol{\mu}_i^*)$ be the unique Lagrange multiplier vectors corresponding to the first order necessary conditions of (P_1) and (P_2), respectively. Then $\psi^*=\boldsymbol{\mu}^*$.□

Assume without loss of generality that $\boldsymbol{B}(\mathbf{x}^*)=\{1,2,\ldots,N_1\}$ for some $N_1\leq N$. The next result characterizes the null space of the Jacobian of the active constraints, namely the matrix $[\nabla\mathbf{g}_1(\mathbf{x}^*),\ldots,\nabla\mathbf{g}_{N_1}(\mathbf{x}^*),\mathbf{L}']$ and its transpose.

Proposition 3.4: Let Assumptions 2.1 and 2.2 hold. The nullspaces of the matrices $[\nabla\mathbf{g}_1(\mathbf{x}^*),\ldots,\nabla\mathbf{g}_{N_1}(\mathbf{x}^*),\mathbf{L}']$ and its transpose are given by $\{(\mathbf{0}',\mathbf{v}')'\mid\mathbf{v}\in\text{Null}(\mathbf{L}')\}$ and $\left\{(\mathbb{1}\otimes v)'\mid v\in\text{Null}\left(\left[\nabla g_1(x^*),\ldots,\nabla g_{N_1}(x^*)\right]'\right)\right\}$, respectively. □

IV. An equivalent optimization problem with equality constraints

In this section we re-formulate (P_2) so that the resulting problem contains only equality constraints, by introducing additional variables. By applying a first order method for solving the first order necessary conditions of this new optimization problem, we in fact derive a distributed algorithm for solving (P_1). Consider the following constrained optimization problem:

$$(P_3)\qquad \min_{\mathbf{x},\mathbf{z}}\quad \mathbf{F}(\mathbf{x}), \tag{7}$$

$$\text{subject to:}\quad \mathbf{h}(\mathbf{x})=\mathbf{L}\mathbf{x}=0, \tag{8}$$

$$\mathbf{g}_i(\mathbf{x})+\mathbf{z}_i^2=0. \tag{9}$$

Remark 4.1: Following a similar avenue as in Section 3.3.2, page 286 of [1], it can be shown that if $(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ is a pair of local minimizer-Lagrange multipliers of (P_2), (with $\boldsymbol{\lambda}^*\in\text{Range}(\mathbf{L})$) then $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ is a pair of local minimizer-Lagrange multipliers for (P_3). In particular, $\mathbf{z}_i^*=-\mathbf{g}_i(\mathbf{x}^*)^{1/2}$.□

Let us now define the augmented Lagrangian for (P_3), namely

$$\begin{aligned}\tilde{\mathcal{L}}_c(\mathbf{x},\mathbf{z},\boldsymbol{\mu},\boldsymbol{\lambda}) \triangleq\ & \mathbf{F}(\mathbf{x})+\sum_{i=1}^N\boldsymbol{\mu}_i\left(\mathbf{g}_i(\mathbf{x})+\mathbf{z}_i^2\right)+\boldsymbol{\lambda}'\mathbf{L}\mathbf{x}+\\ &+\ \frac{c}{2}\sum_i^N\left(\mathbf{g}_i(\mathbf{x})+\mathbf{z}_i^2\right)^2+\frac{c}{2}\mathbf{x}'\mathbf{L}\mathbf{x},\end{aligned} \tag{10}$$

where we sometimes denote $\tilde{\mathcal{L}}_c(\cdot)$ for $c=0$ by $\tilde{\mathcal{L}}_0(\cdot)$. For notational simplification, let us introduce the notations $\tilde{\mathbf{x}}'=(\mathbf{x}',\mathbf{z}')$ and let $\tilde{\mathbf{g}}_i(\tilde{\mathbf{x}})\triangleq\mathbf{g}_i(\mathbf{x})+\mathbf{z}_i^2$. The gradient of $\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})$ with respect to $\tilde{\mathbf{x}}$ is given by $\nabla_{\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})'=[\nabla_{\mathbf{x}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})',\nabla_{\mathbf{z}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})']$, where $\nabla_{\mathbf{x}}\tilde{\mathcal{L}}_c(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\lambda})=\nabla\mathbf{F}(\mathbf{x})+\sum_{i=1}^N\boldsymbol{\mu}_i\nabla\mathbf{g}_i(\mathbf{x})+\mathbf{L}'\boldsymbol{\lambda}+c\sum_{i=1}^N\left(\mathbf{g}_i(\mathbf{x})+\mathbf{z}_i^2\right)\nabla\mathbf{g}_i(\mathbf{x})+c\mathbf{L}\mathbf{x}$ and $\nabla_{\mathbf{z}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})'=$

$\left[2\boldsymbol{\mu}_1\mathbf{z}_1 + 2c\mathbf{z}_1\left(\mathbf{g}_1(\mathbf{x})+\mathbf{z}_1^2\right),\ldots,2\boldsymbol{\mu}_N\mathbf{z}_N + 2c\mathbf{z}_N\left(\mathbf{g}_N(\mathbf{x})+\mathbf{z}_N^2\right)\right]$.
The Hessian of $\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})$ with respect to $\tilde{\mathbf{x}}$ is given by

$$\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) = \begin{bmatrix} \nabla^2_{\mathbf{xx}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) & \nabla^2_{\mathbf{xz}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) \\ \nabla^2_{\mathbf{zx}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) & \nabla^2_{\mathbf{zz}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) \end{bmatrix},$$

where

$$\begin{aligned}\nabla^2_{\mathbf{xx}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) &= \nabla^2\mathbf{F}(\mathbf{x}) + \sum_{i=1}^{N}\boldsymbol{\mu}_i\nabla^2\mathbf{g}_i(\mathbf{x}) + c\sum_{i=1}^{N}\nabla\mathbf{g}_i(\mathbf{x})\nabla\mathbf{g}_i(\mathbf{x})' + \\ &+ c\sum_{i=1}^{N}\left(\mathbf{g}_i(\mathbf{x})+z_i^2\right)\nabla^2\mathbf{g}_i(\mathbf{x}) + c\mathbf{L}, \qquad (11)\end{aligned}$$

$$\nabla^2_{\mathbf{xz}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{zx}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda})' = [2c\mathbf{z}_1\nabla\mathbf{g}_1(\mathbf{x}),\ldots,2c\mathbf{z}_N\nabla\mathbf{g}_N(\mathbf{x})]$$

and $\nabla^2_{\mathbf{zz}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}},\boldsymbol{\mu},\boldsymbol{\lambda}) = 2\mathrm{diag}(\boldsymbol{\mu}_i + 2c\mathbf{z}_i^2)$. Note that the Lagrangian Hessians for both problems (P_2) and (P_3) are independent of $\boldsymbol{\lambda}$.

In the following a correspondence between the properties of the Hessians of problems (P_2) and (P_3) Lagrangian functions is stated.

Proposition 4.1: Let Assumptions 2.1 and 2.2 and let $(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ be a local minimizer-Lagrange multipliers pair of (P_2) and assume that the Hessian of the Lagrangian of (P_2) satisfies the property: $\mathbf{y}'\nabla^2_{\mathbf{xx}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\mathbf{y} > 0$ for all $\mathbf{y}$ such that $\mathbf{Ly} = 0$ and $\nabla\mathbf{g}_i(\mathbf{x}^*)'\mathbf{y} = 0$, with $i \in \boldsymbol{B}(\mathbf{x}^*)$. Then for the corresponding local minimizer-Lagrange multipliers pair $(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ of (P_3), with $\tilde{\mathbf{x}}^* = (\mathbf{x}^*,\mathbf{z}^*)$, the Hessian of the Lagrangian of (P_3) satisfies the property: $\tilde{\mathbf{y}}'\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_0(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\tilde{\mathbf{y}} > 0$ for all $\tilde{\mathbf{y}}$ such that $[\mathbf{L},\mathbf{0}]\tilde{\mathbf{y}} = 0$ and $\nabla\tilde{\mathbf{g}}_i(\tilde{\mathbf{x}}^*)'\tilde{\mathbf{y}} = 0$ for all i.□
The following proposition, which is an adaptation of a result in [4], states that if the Hessian of the Lagrangian satisfies a property as in Proposition 4.1, the Hessian of the augmented Lagrangian of (P_3) can be made positive definite at the local minimizer-Lagrange multiplier pair for large enough c.

Proposition 4.2 ([4]): Let Assumptions 2.1 and 2.2 and let $(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ be a local minimizer-Lagrange multipliers pair of (P_3) and assume that $\tilde{\mathbf{y}}'\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_0(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\tilde{\mathbf{y}} > 0$ for all $\tilde{\mathbf{y}}$ such that $[\mathbf{L},\mathbf{0}]\tilde{\mathbf{y}} = 0$ and $\nabla\tilde{\mathbf{g}}_i(\tilde{\mathbf{x}}^*)'\tilde{\mathbf{y}} = 0$ Then there exists a positive scalar $\bar{c}$, such that the Hessian $\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) \succ 0$ for all $c \geq \bar{c}$.□

The first order necessary conditions for (P_3) expressed in terms of the augmented Lagrangian can be expressed as:

$$\begin{aligned}\nabla_{\mathbf{x}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) &= 0, &(12)\\ \nabla_{\mathbf{z}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) &= 0, &(13)\\ \mathbf{g}_i(\mathbf{x}^*) + \mathbf{z}_i^{*2} &= 0,\ \forall i &(14)\\ \mathbf{Lx}^* &= 0. &(15)\end{aligned}$$

A first order iterative algorithm that can be attempted to solving the necessary conditions, denoted henceforth as `Algorithm 1`, takes the following form:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - \alpha\nabla_{\mathbf{x}}\tilde{\mathcal{L}}_c(\mathbf{x_k},\mathbf{z_k},\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k), &(16)\\ \mathbf{z}_{k+1} &= \mathbf{z}_k - \alpha\nabla_{\mathbf{z}}\tilde{\mathcal{L}}_c(\mathbf{x_k},\mathbf{z_k},\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k), &(17)\\ \boldsymbol{\mu}_{i,k+1} &= \boldsymbol{\mu}_{i,k} + \alpha\left(\mathbf{g}_i(\mathbf{x}_k)+\mathbf{z}_{i,k}^2\right), \forall i &(18)\\ \boldsymbol{\lambda}_{k+1} &= \boldsymbol{\lambda}_k + \alpha\mathbf{L}\mathbf{x}_k. &(19)\end{aligned}$$

for some positive scalar α small enough.

V. Convergence results for `Algorithm 1`

In this section we give conditions under which `Algorithm 1` converges to a local minimizer. The convergence results are based on the spectral properties of a particular matrix; spectral properties presented in the following lemma.

Lemma 5.1: Let Assumptions 2.1 and 2.2 hold and let $(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ with $\tilde{\mathbf{x}}^* = (\mathbf{x}^*,\mathbf{z}^*)$ and $\boldsymbol{\lambda}^* \in \mathrm{Range}(\mathbf{L})$, be a local minimizer-Lagrange multipliers pair of (P_3). In addition, assume that $\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) \succ 0$ for some $c > 0$. Then the eigenvalues of the matrix $\mathbf{M}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ have negative real parts, where

$$\mathbf{M}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) = \begin{bmatrix} \nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*) & \nabla\tilde{\mathbf{g}}(\tilde{\mathbf{x}}^*) & \tilde{\mathbf{L}}' \\ -\nabla\tilde{\mathbf{g}}(\tilde{\mathbf{x}}^*)' & \mathbf{0} & \mathbf{0} \\ -\tilde{\mathbf{L}} & \mathbf{0} & \frac{1}{\alpha}\mathbf{J} \end{bmatrix} \qquad (20)$$

with

$$\nabla\tilde{\mathbf{g}}(\tilde{\mathbf{x}}^*) = \begin{bmatrix} \nabla\mathbf{g}_1(\mathbf{x}^*) & \cdots & \nabla\mathbf{g}_N(\mathbf{x}^*) \\ 2\mathbf{z}_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 2\mathbf{z}_N^* \end{bmatrix},$$

and $\tilde{\mathbf{L}} = [\mathbf{L}\ \mathbf{0}]$, and where $\mathbf{J}$ is the projection operator on $\mathrm{Null}(\mathbf{L}')$ and α is a positive scalar.

Proof: Let β be an eigenvalue of $\mathbf{M}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ and let $(\mathbf{u}',\mathbf{v}',\mathbf{w}')' \neq 0$ be the corresponding eigenvector, where $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are complex vectors of appropriate dimensions. Denoting by $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ the conjugates of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, respectively, we have

$$\mathrm{Re}(\beta)\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2\right) = \mathrm{Re}\left\{\hat{\mathbf{u}}'\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\mathbf{u} + \hat{\mathbf{w}}'\frac{1}{\alpha}\mathbf{J}\mathbf{w}\right\}.$$

Since $\mathbf{J}$ is a semi-positive definite matrix and $\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ is positive definite (as per our assumption), we have that $\mathrm{Re}(\beta)\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2\right) > 0$, as long as $\mathbf{u} \neq 0$ or $\mathbf{w} \notin \mathrm{Range}(\mathbf{L})$ and therefore $\mathrm{Re}(\beta) > 0$. In the case $\mathbf{u} = 0$ and $\mathbf{w} \in \mathrm{Range}(\mathbf{L})$, we get

$$\mathbf{M}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)[0,\mathbf{v},\mathbf{w}]' = \beta[0,\mathbf{v},\mathbf{w}]',$$

from where we obtain

$$\nabla\tilde{\mathbf{g}}(\mathbf{x}^*)\mathbf{v} + \mathbf{L}'\mathbf{w} = 0,$$

which can be explicitly written as

$$[\nabla\mathbf{g}_1(\mathbf{x}^*),\ldots,\nabla\mathbf{g}_N(\mathbf{x}^*)]\,\mathbf{v} + \mathbf{L}'\mathbf{w} = 0 \qquad (21)$$

and

$$\begin{bmatrix} 2\mathbf{z}_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 2\mathbf{z}_N^* \end{bmatrix}\mathbf{v} = 0. \qquad (22)$$

Since $\mathbf{z}_i^* = -\mathbf{g}_i(\mathbf{x}^*)^{1/2}$ it follows that $\mathbf{z}_i^* > 0$ for all $i \notin \boldsymbol{B}(\mathbf{x}^*)$ and hence, from (22) we have that $\mathbf{v}_i = 0$ for all $i \notin \boldsymbol{B}(\mathbf{x}^*)$. Assume without loss of generality that $\boldsymbol{B}(\mathbf{x}^*) = \{1,2,\ldots,N_1\}$. Equation (21) becomes $\left[\nabla\mathbf{g}_1(\mathbf{x}^*),\ldots,\nabla\mathbf{g}_{N_1}(\mathbf{x}^*)\right]\left[\mathbf{v}_1,\ldots,\mathbf{v}_{N_1}\right]' + \mathbf{L}'\mathbf{w} = 0$. But from Proposition 3.4, we have that $\mathbf{v} = 0$ and $\mathbf{w} \in \mathrm{Null}(\mathbf{L}')$ and since $\mathbf{w} \in \mathrm{Range}(\mathbf{L})$ as well, it must be that $\mathbf{w} = 0$. Hence we have a contradiction since we assumed that $(\mathbf{u}',\mathbf{v}',\mathbf{w}') \neq \mathbf{0}'$ and therefore the real part of β must be positive. ■

We are now ready to present conditions under which local convergence is achieved by iteration (16)-(19).

Theorem 5.1: Let Assumptions 2.1 and 2.2 hold and let $(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ with $\tilde{\mathbf{x}}=(\mathbf{x}^*,\mathbf{z}^*)$, and $\boldsymbol{\lambda}^*\in\text{Range}(\mathbf{L})$, be a local minimizer-Lagrange multipliers pair of (P_3). Assume also that $\tilde{\mathbf{y}}'\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_0(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\tilde{\mathbf{y}}>0$ for all $\tilde{\mathbf{y}}$ such that $[\mathbf{L},\mathbf{0}]\tilde{\mathbf{y}}=0$ and $\nabla\tilde{\mathbf{g}}_i(\tilde{\mathbf{x}}^*)'\tilde{\mathbf{y}}=0$. Then there exists $\bar{c}>0$ so that for all $c>\bar{c}$ we can find $\bar{\alpha}(c)$ such that for all $\alpha\in(0,\bar{\alpha}(c)]$, the set $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*+\text{Null}(\mathbf{L}'))$ is an attractor of iteration (16)-(19). In addition, if the sequence $\{\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k\}$ converges to the set $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*+\text{Null}(\mathbf{L}'))$, the rate of convergence of $\|\mathbf{x}_k-\mathbf{x}^*\|$, $\|\mathbf{z}_k-\mathbf{z}^*\|$, $\|\boldsymbol{\mu}_k-\boldsymbol{\mu}^*\|$ and $\|\boldsymbol{\lambda}_k-[\boldsymbol{\lambda}^*+\text{Null}(\mathbf{S}')]\|$ is linear.

Proof: Using the Lagrangian function defined in (15), iteration (16)-(19) can be equivalently expressed as

$$\begin{bmatrix}\mathbf{x}_{k+1}\\ \mathbf{z}_{k+1}\\ \boldsymbol{\mu}_{k+1}\\ \boldsymbol{\lambda}_{k+1}\end{bmatrix}=\underbrace{\begin{bmatrix}\mathbf{x}_k-\alpha\nabla_{\mathbf{x}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)\\ \mathbf{z}_k-\alpha\nabla_{\mathbf{z}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)\\ \boldsymbol{\mu}_k+\alpha\nabla_{\boldsymbol{\mu}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)\\ \boldsymbol{\lambda}_k+\alpha\nabla_{\boldsymbol{\lambda}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)\end{bmatrix}}_{\mathbf{T}(\alpha,\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)} \quad (23)$$

It can be easily checked that $(\mathbf{x}^*,\mathbf{z}^*\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*+\text{Null}(\mathbf{L}'))$ is a set of fixed points of $\mathbf{T}$. Let us now consider the transformation $\tilde{\boldsymbol{\lambda}}=(\mathbf{I}-\mathbf{J})\boldsymbol{\lambda}$, where $\mathbf{J}$ is the orthogonal projection operator on $\text{Null}(\mathbf{L}')$. This transformation extracts the projection of $\boldsymbol{\lambda}$ on the nullspace of $\mathbf{L}'$ from $\boldsymbol{\lambda}$ and therefore $\tilde{\boldsymbol{\lambda}}$ is the error between $\boldsymbol{\lambda}$ and its orthogonal projection on $\text{Null}(\mathbf{L}')$. Under this transformation, iteration (23) becomes

$$\begin{bmatrix}\mathbf{x}_{k+1}\\ \mathbf{z}_{k+1}\\ \boldsymbol{\mu}_{k+1}\\ \tilde{\boldsymbol{\lambda}}_{k+1}\end{bmatrix}=\underbrace{\begin{bmatrix}\mathbf{x}_k-\alpha\nabla_{\mathbf{x}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k)\\ \mathbf{z}_k-\alpha\nabla_{\mathbf{z}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k)\\ \boldsymbol{\mu}_k+\alpha\nabla_{\boldsymbol{\mu}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k)\\ (\mathbf{I}-\mathbf{J})\tilde{\boldsymbol{\lambda}}_k+\alpha\nabla_{\boldsymbol{\lambda}}\mathcal{L}_c(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k)\end{bmatrix}}_{\tilde{\mathbf{T}}(\alpha,\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k)}$$

where we used the fact that $(\mathbf{I}-\mathbf{J})\tilde{\boldsymbol{\lambda}}=(\mathbf{I}-\mathbf{J})\boldsymbol{\lambda}$, $(\mathbf{I}-\mathbf{J})\mathbf{L}\mathbf{x}=\mathbf{L}\mathbf{x}$, since $\mathbf{L}\mathbf{x}\in\text{Range}(\mathbf{L})$, and $\mathbf{L}'\boldsymbol{\lambda}=\mathbf{L}'(\tilde{\boldsymbol{\lambda}}+\mathbf{J}\boldsymbol{\lambda})=\mathbf{L}'\tilde{\boldsymbol{\lambda}}$. Clearly $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ is a fixed point for $\tilde{\mathbf{T}}$ and if $\left(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k\right)$ converges to $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$, this implies that $(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\boldsymbol{\lambda}_k)$ converges to $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*+\text{Null}(\mathbf{L}'))$. The derivative of the mapping $\tilde{\mathbf{T}}\left(\alpha,\mathbf{x},\mathbf{z},\boldsymbol{\mu},\tilde{\boldsymbol{\lambda}}\right)$ at $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ is given by $\nabla\tilde{\mathbf{T}}(\alpha,\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)=\mathbf{I}-\alpha\mathbf{M}_c(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$, where $\mathbf{M}_c(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ was defined in (20). Proposition 4.2 tells us that there exists a $\bar{c}$ so that $\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_c(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\succ 0$ for all $c\geq\bar{c}$. Therefore, for all such c, by Lemma 5.1 we have that the real parts of the eigenvalues of $\mathbf{M}_c(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ are positive. Consequently, using a continuity argument, we can find an $\bar{\alpha}$ (which depends on c) so that for all $\alpha\in(0.\bar{\alpha}(c)]$, the eigenvalues of $\nabla\tilde{\mathbf{T}}(\alpha,\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ are strictly within the unit circle. Using a similar argument as in Proposition 4.4.1, page 387, [1], there exist a norm $\|\cdot\|$ and a sphere $\mathcal{S}_\epsilon=\left\{(\mathbf{x}',\mathbf{z}',\boldsymbol{\mu}',\boldsymbol{\lambda}')'\mid\|(\mathbf{x}',\mathbf{z}',\boldsymbol{\mu}',\boldsymbol{\lambda}')'-(\mathbf{x}^{*\prime},\mathbf{z}^{*\prime},\boldsymbol{\mu}^{*\prime},\boldsymbol{\lambda}^{*\prime})'\|<\epsilon\right\}$ for some $\epsilon>0$ so that the induced norm of $\nabla\tilde{\mathbf{T}}(\alpha,\mathbf{x},\mathbf{z},\boldsymbol{\mu},\boldsymbol{\lambda})$ is less than one within the sphere $\mathcal{S}_\epsilon$. Therefore, using the mean value theorem, it follows that $\tilde{\mathbf{T}}(\alpha,\mathbf{x},\mathbf{z},\boldsymbol{\mu},\boldsymbol{\lambda})$ is a contraction map for any vector in the sphere $\mathcal{S}_\epsilon$. By invoking the contraction map theorem (see for example Chapter 7 of [5]) we get that $\left(\mathbf{x}_k,\mathbf{z}_k,\boldsymbol{\mu}_k,\tilde{\boldsymbol{\lambda}}_k\right)$ converges to $(\mathbf{x}^*,\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)$ for any initial value in S_ϵ. ■

VI. Distributed algorithm

In what follows we connect the convergence results for (P_3) (and consequently (P_2)) to the solution of (P_1). First let us re-write `Algorithm 1` by emphasizing each i^{th} n-dimensional component of $\mathbf{x}_k$. As a result, we obtain the following iteration, that we refer to as `Algorithm 2`:

$$\begin{aligned}x_{i,k+1} &= x_{i,k}-\alpha\Big\{\nabla f_i(x_{i,k})+\mu_{i,k}\nabla g_i(x_{i,k})\\ &+\sum_{j\in\mathcal{N}_i}l_{ij}(\lambda_{i,k}-\lambda_{j,k})+c\left(g_i(x_{i,k})+z_{i,k}^2\right)\nabla g_i(x_{i,k})\\ &+c\sum_{j\in\mathcal{N}_i}l_{ij}(x_{i,k}-x_{j,k})\Big\} \quad (24)\\ z_{i,k+1} &= z_{i,k}-2\alpha\left\{\mu_{i,k}z_{i,k}+cz_{i,k}\left(g_i(x_{i,k})+z_{i,k}^2\right)\right\} \quad (25)\\ \mu_{i,k+1} &= \mu_{i,k}+\alpha\left(g_i(x_{i,k})+z_{i,k}^2\right) \quad (26)\\ \lambda_{i,k+1} &= \lambda_{i,k}+\alpha\sum_{j\in\mathcal{N}_i}l_{ij}(x_{i,k}-x_{j,k}) \quad (27)\end{aligned}$$

In the above iteration we used the fact that L is assumed symmetric and therefore $\mathbf{L}'\boldsymbol{\lambda}=\mathbf{L}\boldsymbol{\lambda}$. It can be easily observed that the algorithm is distributed (assuming the scalars α and c are globally known). Indeed iteration (24)-(27) corresponds to the operations each agent i executes at each time instant; operations for which only local information, or information from neighbors is used. The following corollary gives the conditions under which `Algorithm 2` ensures convergence to a local minimizer of (P_1).

Corollary 6.1: Let Assumptions 2.1 and 2.2 hold and let (x^*,ψ^*) be a local minimizer-Lagrange multiplier pair of (P_1). Assume also that $y'\sum_{i=1}^N\left[\nabla^2 f_i(x^*)+\psi_i^*\nabla^2 g_i(x^*)\right]y>0$ for all y such that $\nabla g_i(x^*)'y=0$, with $i\in B(x^*)$. Then there exists $\bar{c}>0$ so that for all $c\geq\bar{c}$ we can find $\bar{\alpha}(c)$ such that for all $\alpha\in(0,\bar{\alpha}(c)]$, (x^*,ψ^*) is a point of attraction for iteration (24)-(27), for all $i=1,\dots,N$. In addition, if the sequence $\{x_{i,k},\mu_{i,k}\}$ converges to (x^*,ψ^*), then the rate of convergence of $\|x_{i,k}-x^*\|$ and $\|\mu_{i,k}-\psi^*\|$ is linear.

Proof: By Proposition 3.2 we have that $\mathbf{x}^*=\mathbb{1}\otimes x^*$ is a local minimizer of (P_2) with corresponding Lagrange multipliers $(\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*+\text{Null}(\mathbf{L}'))$, with $\boldsymbol{\lambda}^*\in\text{Range}(\mathbf{L})$. In addition, by Proposition 3.3 we have that $\boldsymbol{\mu}^*=\psi^*$. Using the definition of the Lagrangian function of (P_2) introduced in (5), we have

$$\nabla^2_{\mathbf{x}\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)=\text{diag}\left(\nabla^2 f_i(x^*)+\psi_i^*\nabla^2 g_i(x^*),i=1,\dots,N\right).$$

Recalling the assumption that $y'\sum_{i=1}^N\left[\nabla^2 f_i(x^*)+\psi_i^*\nabla^2 g_i(x^*)\right]y>0$ for all y such that $\nabla g_i(x^*)'y=0$, with $i\in B(x^*)$, by Proposition 3.4 this is equivalent to $\mathbf{y}'\nabla^2_{\mathbf{x}\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\mathbf{y}>0$ for all $\mathbf{y}$ such that $\mathbf{L}\mathbf{y}=0$ and $\nabla\mathbf{g}_i(\mathbf{x}^*)'\mathbf{y}=0$, with $i\in\boldsymbol{B}(\mathbf{x}^*)$. Furthermore, applying Proposition 4.1 we have that the Hessian of the Lagrangian of (P_3) satisfies $\tilde{\mathbf{y}}'\nabla^2_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\tilde{\mathcal{L}}_0(\tilde{\mathbf{x}}^*,\boldsymbol{\mu}^*,\boldsymbol{\lambda}^*)\tilde{\mathbf{y}}>0$ for all $\tilde{\mathbf{y}}$ such that $[\mathbf{L},\mathbf{0}]\tilde{\mathbf{y}}=0$ and $\nabla\tilde{\mathbf{g}}_i(\tilde{\mathbf{x}}^*)'\tilde{\mathbf{y}}=0$ for all i. All conditions of Theorem 5.1 are satisfied, and the result follows. ■

A. Distributed algorithm for mixed constraints

In this section we show how the previous algorithm can be extended to the case where each agent has also an equality constraint in addition to the inequality one. Consider the following optimization problem:

$$
\begin{aligned}
(P_1') \quad & \min_{x\in\mathbb{R}^n} && \sum_{i=1}^{N} f_i(x), \\
& \text{subject to:} && g_i(x) \leq 0, i = 1, \ldots, N \\
& && \xi_i(x) = 0, i = 1, \ldots, N,
\end{aligned}
$$

where Assumptions 2.1 and 2.2 *are extended* to include the (twice) differentiability of $\xi_i(x)$ and the linear independence of the set of vectors $\{\nabla \xi_j(x^*)\}_{j=1}^N$ and $\{\nabla g_i(x^*)\}_{i\in B(x^*)}$. Consider now the following iteration, referenced henceforth as `Algorithm 3`:

$$
\begin{aligned}
x_{i,k+1} &= x_{i,k} - \alpha\Big\{\nabla f_i(x_{i,k}) + \mu_{i,k}\nabla g_i(x_{i,k}) + \eta_{i,k}\nabla\xi_i(x_{i,k}) \\
&+ \sum_{j\in\mathcal{N}_i} l_{ij}(\lambda_{i,k}-\lambda_{j,k}) + c\left(g_i(x_{i,k}) + z_{i,k}^2\right)\nabla g_i(x_{i,k}) \\
&+ c\xi_i(x_{i,k})\nabla\xi_i(x_{i,k}) + c\sum_{j\in\mathcal{N}_i} l_{ij}(x_{i,k}-x_{j,k})\Big\} \qquad (28) \\
z_{i,k+1} &= z_{i,k} - 2\alpha\left\{\mu_{i,k} z_{i,k} + c z_{i,k}\left(g_i(x_{i,k}) + z_{i,k}^2\right)\right\} \qquad (29) \\
\mu_{i,k+1} &= \mu_{i,k} + \alpha\left(g_i(x_{i,k}) + z_{i,k}^2\right) \qquad (30) \\
\lambda_{i,k+1} &= \lambda_{i,k} + \alpha\sum_{j\in\mathcal{N}_i} l_{ij}(x_{i,k}-x_{j,k}) \qquad (31) \\
\eta_{i,k+1} &= \eta_{i,k} + \alpha\xi_i(x_{i,k}) \qquad (32)
\end{aligned}
$$

The following result states the convergence properties of this algorithm.

Proposition 6.1: Let the extended Assumptions 2.1 and 2.2 hold and let (x^*, ψ^*, η^*) be a local minimizer-Lagrange multiplier pair of (P_1'). Assume also that $y'\sum_{i=1}^N\left[\nabla^2 f_i(x^*) + \psi_i^*\nabla^2 g_i(x^*) + \eta_i^*\nabla^2\xi_i(x^*)\right]y > 0$ for all y such that $\nabla g_i(x^*)'y = 0$, with $i \in B(x^*)$ and $\nabla\xi_i(x^*)'y = 0$ for all i. Then there exists $\bar{c} > 0$ so that for all $c \geq \bar{c}$ we can find $\bar{\alpha}(c)$ such that for all $\alpha \in (0, \bar{\alpha}(c)]$, (x^*, ψ^*, η^*) is a point of attraction for iteration (28)-(32), for all $i = 1, \ldots, N$. In addition, if the sequence $\{x_{i,k}, \mu_{i,k}, \eta_{i,k}\}$ converges to (x^*, ψ^*, η^*), then the rate of convergence of $\|x_{i,k} - x^*\|$, $\|\mu_{i,k} - \psi^*\|$ and $\|\eta_{i,k} - \eta^*\|$ is linear.□

The proof of this result would follow the same steps as in the case of Corollary 6.1. Note that as in the case of `Algorithm 2`, `Algorithm 3` is distributed as well since the agents update their estimates of the minimizer and Lagrange multipliers using only local information and information from neighbors.

VII. Conclusions

We presented a distributed algorithm for solving an optimization problem with inequality constrains; where the cost function is expressed as a sum of functions and each agent is aware of only one function of the sum and has its own local inequality constraint. We gave conditions for the (local) convergence of the algorithm where special care had to be paid to dealing with the non-regularity of the local minimizers due to the distributed formulation. In addition, we introduced an extension of the algorithm that solves the distributed optimization problem in the case agents are endowed with equality constraints as well, in addition to the inequality ones.

References

[1] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 1999.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.

[3] A. Cherukuri and J. Cortés. Distributed generator coordination for initialization and anytime optimization in economic dispatch. *IEEE Transactions on Control of Network Systems (Submitted)*, 2013.

[4] M.R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[5] V.I. Istratescu. *Fixed Point Theory, An Introduction*. D.Reidel, the Netherlands., 1981.

[6] D. Jakovetic, M.F.J. Moura, and J. Xavier. Linear convergence rate of a class of distributed augmented lagrangian algorithms. *CoRR*, abs/1307.2482, 2013.

[7] D. Jakovetic, J. Xavier, and J.M.F. Moura. Cooperative convex optimization in networked systems: Augmented lagrangian algorithms with directed gossip communication. *Signal Processing, IEEE Transactions on*, 59(8):3889–3902, Aug.

[8] B. Johansson, T. Keviczky, M. Johansson, and K.H. Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems. *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 4185–4190, Dec 2008.

[9] B. Johansson, M. Rabi, and K.H. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2009.

[10] I. Lobel and A. Ozdaglar. Distributed subgradient methods for convex optimization over random networks. *Automatic Control, IEEE Transactions on*, 56(6):1291–1306, june 2011.

[11] I. Lobel, A. Ozdaglar, and D. Feijer. Distributed multi-agent optimization with state-dependent communication. *Mathematical Programming*, 129(2):255–284, 2011.

[12] I. Matei and J.S. Baras. Performance evaluation of the consensus-based distributed subgradient method under random communication topologies. *Selected Topics in Signal Processing, IEEE Journal of*, 5(4):754–771, aug. 2011.

[13] I. Matei and J.S. Baras. Distributed algorithms for optimization problems with equality constraints. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 2352–2357, Dec 2013.

[14] I. Matei, J.S. Baras, M. Nabi, and T. Kurtoglu. An extension of the method of multipliers for distributed nonlinear programming. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 6951–6956, Dec 2014.

[15] A. Nedic. Asynchronous broadcast-based convex optimization over a network. *Automatic Control, IEEE Transactions on*, 56(6):1337 –1351, june 2011.

[16] A. Nedic, A. Ozdaglar, and P.A. Parrilo. Constrained consensus and optimization in multi-agent networks. *Automatic Control, IEEE Transactions on*, 55(4):922 –938, april 2010.

[17] A. Nedic and A. Ozdalgar. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control*, 54(1):48–61, Jan 2009.

[18] S. Sundhar Ram, A. Nedic, and V. V. Veeravalli. Distributed stochastic subgradient projection algorithms for convex optimization. *Journal of Optimization Theory and Applications*, pages 516–545, 2010.

[19] W. Shi, Q. Ling, K. Yuan, G Wu, and W. Yin. On the linear convergence of the admm in decentralized consensus optimization. *Signal Processing, IEEE Transactions on*, 62(7):1750–1761, April 2014.

[20] K. Srivastava and A. Nedic. Distributed asynchronous constrained stochastic optimization. *Selected Topics in Signal Processing, IEEE Journal of*, 5(4):772 –790, aug. 2011.

[21] H. Terelius, U. Topcu, and R. M. Murray. Decentralized multi-agent optimization via dual decomposition. In *18th World Congress of the International Federation of Automatic Control (IFAC)*, August 2011.

[22] E. Wei and A. Ozdaglar. Distributed alternating direction method of multipliers. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 5445–5450, Dec 2012.

[23] M. Zhu and S. Martinez. An approximate dual subgradient algorithm for multi-agent non-convex optimization. *Automatic Control, IEEE Transactions on*, 58(6):1534–1539, June 2013.