Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-2 (2020) 5417–5422

A Newton consensus method for distributed optimization ⋆

Martin Guay ∗

∗ *Queen's University, Department of Chemical Engineering, Kingston, Ontario, Canada (e-mail: guaym@queensu.ca)*

Abstract: This manuscript proposes a distributed Newton seeking for the solution of distributed optimization problems with locally measured but unknown cost functions. The approach implements a Newton step for both the primal and dual problems that can be implemented in a completely decentralized fashion. Unlike existing techniques, no exchange of derivative information between agents is required. In addition, no explicit inversion of the Hessian information is required to generate the required Newton step. The local gradients and Hessians are estimated using a perturbation based extremum seeking control technique. A simulation study demonstrates the effectiveness of the technique.

Keywords: Newton consensus, extremum seeking, distributed optimization.

1. INTRODUCTION

The last decade has seen significant advances in the development of distributed and decentralized optimization techniques for large-scale problems operating over communication networks. The design of distributed optimization algorithm is now an established problem in the literature. One of the most studied problem is the consensus optimization problem (*e.g.* (Bertsekas and Tsitsiklis, 1989), (Nedić and Ozdaglar, 2009), (Johansson et al., 2009)). In this problem, agents operating over a communication network are made to achieve consensus at the optimum of a total network cost. In the standard framework for this problem, the agents have access to measurements of its local costs and they can communicate local decision variables to neighboring agents.

Two basic classes of techniques are used to solve this class of problems. The subgradient based techniques such as (Nedić and Ozdaglar, 2009) are usually recognized to yield simple algorithms with typically slow convergence. The Alternating Direction Method of Multipliers (ADMM) based techniques ((Boyd et al., 2011), (Schizas et al., 2008)) decompose the consensus optimization problem in two alternating optimization problems that provide updates of dual variables. Newton consensus methods (Wei et al. (2013), Zanella et al. (2011)) have been shown to provide significant performance improvements over subgradient techniques in consensus optimization. However, they have not been shown to outperform ADMM. In recent developments, a novel Newton consensus method was proposed in (Tutunov et al., 2019). This technique enables the approximation of the Newton step for the consensus problem without the need to exchange local derivative information over the communication network.

The objective of this study is to develop a model-free distributed Newton seeking that implements a similar distributed Newton consensus. The approach combines a local extremum seeking control (ESC) approach for each agent and a Newton consensus approach inspired by the dual approach recently proposed by Tutunov et al. (2019). The proposed ESC technique for the local optimization problems is based on the Newton seeking approach of Ghaffari et al. (2012). ESC has demonstrated considerable potential (Xu and Soh (2013),Michalowsky et al. (2018), Guay et al. (2018)) for the design of distributed optimization and control systems. Most existing technique rely on gradient based consensus algorithms which suffer from the same performance limitations of subgradient techniques. The main contribution of this study is the design of a continuous time Newton consensus based on the methodology presented in (Tutunov et al., 2019). The combination with the local Newton seeking algorithm enables the implementation of a network wide Newton consensus optimization using a model-free approach.

The paper is organized as follows. The problem statement is presented in Section 2 along with some preliminaries and a brief description of the dual problem. Section 3 presents the proposed Newton consensus technique. A brief stability analysis is performed to highlights the main properties of the Newton consensus. A detailed proof is outside the scope of the current study and will be presented in future work. The distributed Newton seeking approach is given in Section 5. A simulation study is provided in Section 6 followed by brief conclusions in Section 7.

2. PROBLEM DEFINITION

The objective of this study is the solution of the optimization problem:

$$\min_{x_1,\ldots,x_m} \sum_{i=1}^{m} f_i(x_i), \quad \text{s.t. } x_1 = \cdots = x_m \tag{1}$$

⋆ The author acknowledges the financial support of the Natural Sciences and Engineering Research Council of Canada and the Fields Institute.

10.1016/j.ifacol.2020.12.1536

where $x_i \in \mathbb{R}^n$, and each function $f_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable functions for $i = 1, \ldots, m$.

Assumption 1. The Hessian of function f_i are positive definite constant matrices and

$$\alpha_1 I \leq \frac{\partial f_i}{\partial x_i \partial x_i^T} \leq \alpha_2 I$$

for some positive constants α_1 and α_2 $\forall x_i \in \mathbb{R}^n$, for $i = 1, \ldots, m$.

Assumption 2. The total objective function

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

has a unique minimum at $x = x^*$ such that

$$\left.\frac{\partial f}{\partial x}\right|_{x=x^*} = 0.$$

Problem (1) is solved over a communication network of m agents modeled as an undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ denotes the set of vertices of the graph with $|\mathcal{V}| = m$ while $\mathcal{E}$ denotes the set of edges of $\mathcal{G}$ with $|\mathcal{E}| = p$. Each vertex corresponds to an agent who is allowed to communication with its immediate neighbors along the edges of $\mathcal{G}$. The neighborhood of agent i is denoted by $\mathcal{N}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}\}$. The degree of agent i, d_i, is given by the number of neighbors in $\mathcal{N}_i$, $|\mathcal{N}_i| = d_i$.

The structure of the graph can be summarized by its Laplacian, $\mathcal{L} = D - A$ where $D \in \mathbb{R}^{m\times m}$ is the degree matrix and $A \in \mathbb{R}^{m\times m}$ is the adjency matrix. The adjacency matrix has elements $a_{ij} = 1$ whenever $(i, j) \in \mathcal{E}$ for $v_i, v_j \in V$ and $a_{ij} = 0$ otherwise. A graph $\mathcal{G}$ is undirected if $(i, j) \in \mathcal{E}$ whenever $(j, i) \in \mathcal{E}$. The degree matrix D is a diagonal matrix where $d_i = \sum_{j=1}^{m} a_{ij}$. The Laplacian matrix is given by $\mathcal{L} = D - A$. Let $\mathbf{1}_m$ be an m dimensional vector of ones. We consider a m-dimensional vector $\boldsymbol{r}$ and a $m \times (m-1)$ matrix $\boldsymbol{R}$ such that:

$$\boldsymbol{r} = \frac{1}{\sqrt{m}}\mathbf{1}_m,\ \boldsymbol{r}^T\boldsymbol{R} = 0,$$

$$\boldsymbol{R}^T\boldsymbol{R} = I_{m-1},\ \boldsymbol{R}\boldsymbol{R}^T = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_n^T.$$

For an undirected graph, the Laplacian matrix is such that:

$$\mathbf{1}_m^T\mathcal{L} = 0,\ \mathcal{L}\mathbf{1}_m = 0.$$

Throughout this paper, it is assumed that all local costs f_i are unknown but available for measurement for agent i. The structure of the network $\mathcal{L}$ is known.

2.1 Dual problem

In this section, we pose the dual problem of (1). Following the approach presented in (Tutunov et al., 2019), we define auxiliary variables, $y_j \in \mathbb{R}^m$, where y_j is the vector of all j^{th} elements of the local agent state x_i, for $i = 1, \ldots, m$. That is,

$$y_j = [x_1(j),\, x_2(j),\, \ldots,\, x_m(j)]^T$$

for $j = 1, \ldots, n$.

Using this notation, the original problem can be written:

$$\min_{y_1,\ldots,y_n} \sum_{i=1}^{m} f_i(y_1(i), \ldots, y_n(i)) \tag{2}$$

$$s.t. \quad \mathcal{L}y_j = 0, \qquad j = 1, \ldots, n. \tag{3}$$

We define the matrix $\mathcal{M} = \mathcal{L} \otimes I_{n\times n}$, an element of $\mathbb{R}^{nm\times nm}$.

The Lagrangian associated with (1) is given by:

$$\begin{aligned} L(y_1, \ldots, y_n, \lambda_1, \ldots, \lambda_n) = &\sum_{i=1}^{m} f_i(y_1(i), \ldots, y_n(i)) \\ &+ \lambda_1^T\mathcal{L}y_1 + \ldots + \lambda_n^T\mathcal{L}y_n. \end{aligned} \tag{4}$$

where $\lambda_i \in \mathbb{R}^m$ are the Lagrange multipliers.

By the properties of the graph Laplacian ($\mathcal{L} = \mathcal{L}^T$), the Lagrangian can be written as follows:

$$\begin{aligned} L(y_1, \ldots, y_n, \lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{m} \Big(&f_i(y_1(i), \ldots, y_n(i)) \\ &+ (\mathcal{L}\lambda_1)_i y_1(i) + \ldots + (\mathcal{L}\lambda_n)_i y_n(i)\Big). \end{aligned} \tag{5}$$

The dual problem associated with (5) is given as follows:

$$\begin{aligned} \boldsymbol{\nu}(\boldsymbol{\lambda}) = \sum_{i=1}^{m} \inf_{y_1,\ldots,y_n} \Big(&f_i(y_1(i), \ldots, y_n(i)) \\ &+ (\mathcal{L}\lambda_1)_i y_1(i) + \ldots + (\mathcal{L}\lambda_n)_i y_n(i)\Big). \end{aligned} \tag{6}$$

The primal problem can be defined as the minimization of the Lagrangian at a fixed value of the Lagrange multipliers. The corresponding primal variables are the corresponding solutions denoted by $\mathbf{y}(\boldsymbol{\lambda})$ where $\mathbf{y} = \left[y_1^T, y_2^T, \ldots, y_n^T\right]^T$ and $\boldsymbol{\lambda} = \left[\lambda_1^T, \lambda_2^T, \ldots, \lambda_n^T\right]^T$.

3. NEWTON CONSENSUS

In the following, we extend the notation used to define $\mathbf{y}$ and $\boldsymbol{\lambda}$ to the following auxiliary variables $\mathbf{z} = \left[z_1^T, z_2^T, \ldots, z_n^T\right]^T$ and $\boldsymbol{\delta} = \left[\delta_1^T, \delta_2^T, \ldots, \delta_n^T\right]^T$.

Using standard arguments Tutunov et al. (2019), the gradient of dual function $\nu(\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$ is given by:

$$\nabla\nu(\boldsymbol{\lambda}) = \mathcal{M}\mathbf{y}(\boldsymbol{\lambda}). \tag{7}$$

Its Hessian is as follows:

$$H(\boldsymbol{\lambda}) = -\mathcal{M}(\nabla^2 f(\mathbf{y}(\boldsymbol{\lambda})))^{-1}\mathcal{M}. \tag{8}$$

The objective of this section is to develop a distributed algorithm that implements a Newton algorithm for the Newton step, $\boldsymbol{\delta}(\boldsymbol{\lambda})$,

$$H(\boldsymbol{\lambda})\boldsymbol{\delta}(\boldsymbol{\lambda}) = \mathcal{M}\mathbf{y}(\boldsymbol{\lambda}). \tag{9}$$

In Tutunov et al. (2019), this task is achieved using an approximate inversion algorithm that exploits the symmetric diagonally dominant property of the matrices $\mathcal{M}$ and $H(\boldsymbol{\lambda})$. In this study, the Newton step is computed implicitly as the unique equilibrium of a dynamical system. The prototype dynamical system proposed in this study is given by:

$$\dot{\mathbf{z}} = -\gamma(\mathcal{M}\mathbf{z} - \mathcal{M}\mathbf{y}(\boldsymbol{\lambda})) \tag{10}$$

$$\dot{\boldsymbol{\delta}} = -\gamma\Big(\mathcal{M}\boldsymbol{\delta} + \nabla^2 f(\mathbf{y}(\boldsymbol{\lambda}))\mathbf{z}\Big) \tag{11}$$

where γ is a positive gain to be assigned.

By the properties of the Laplacian matrix, it follows that the matrix $\mathcal{M}$ is a positive semi-definite symmetric matrix. As a result, the linear dynamical system (10) has a unique exponentially stable equilibrium at $\mathcal{M}\mathbf{z} = \mathcal{M}\mathbf{y}(\boldsymbol{\lambda})$. Similarly, the dynamical system (11) has a unique equilibrium at $\mathcal{M}\boldsymbol{\delta} = -\nabla^2 f(\mathbf{y}(\boldsymbol{\lambda}))\boldsymbol{z}$.

The corresponding equilibrium value of δ is therefore the unique solution of the Newton step equation (9). This Newton step is implemented using the following update of the Lagrange multiplier:

$$\dot{\boldsymbol{\lambda}} = \rho\boldsymbol{\delta} \tag{12}$$

where ρ is a positive constant to be assigned.

3.1 Decentralized Newton consensus

The dynamical systems (10), (11) and (12) can be implemented in a completely distributed fashion requiring the exchange of the neighboring values of the $y_j(k)$,$z_j(k)$, $\delta_j(k)$ and $\lambda_j(k)$ for $k \in \mathcal{N}_i$, with $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

To see this, we expand the elements of (10) and (11) componentwise. Elements of the dynamics of the auxiliary variables, $\mathbf{z}$, are given by:

$$\dot{z}_r(k) = -\gamma\left(\sum_{j=1}^{m}\mathcal{L}_{kj}z_r(j) - \sum_{j=1}^{m}\mathcal{L}_{kj}y_r(j)\right). \tag{13}$$

For the Newton step variables, one obtains:

$$\dot{\delta}_r(k) = -\gamma\left(\sum_{j=1}^{m}\mathcal{L}_{kj}\delta_r(j) + \sum_{j=1}^{n}\frac{\partial^2 f_k}{\partial y_r(k)\partial y_j(k)}z_j(k)\right) \tag{14}$$

which yields the following update of the Lagrange multipliers:

$$\dot{\lambda}_r(k) = \rho\delta_r(k) \tag{15}$$

for $r = 1, \ldots, n$, $k = 1, \ldots, m$.

It follows that the computation of the updates requires neighboring values from z_r, δ_r and y_r. More importantly, it can be seen from (14) that the Newton consensus step for the network can be calculated using only the Hessian of the local cost with respect to the local variables $x_k = [y_1(k), \ldots, y_n(k)]^T$. Moreover, the proposed approach does not require any explicit inversion of the Hessian to compute the Newton step.

3.2 Approximation of the primal variables

Formally, the primal variables must be chosen as the solution of the partial differential equations:

$$\frac{\partial f_i}{\partial y_j(i)} = -(\mathcal{L}\lambda_j)_i. \tag{16}$$

While this is possible for simple quadratic local cost functions, the computation of primal variables can become exceedingly complex in a network application. In this study, we propose a simple approximation that computes the local minimizer of the Lagrangian at a constant value of the Lagrange multiplier vector.

We propose the compute the local minimizer using a Newton method. For agent i, the elements of the gradient of the Lagrangian function are given by:

$$\frac{\partial L}{\partial y_j(i)} = \frac{\partial f_i}{\partial y_j(i)} + (\mathcal{L}\lambda_j)_i = g_j(i) \tag{17}$$

while the elements of the Hessian are of the form:

$$\frac{\partial^2 L}{\partial y_j(i)\partial y_k(i)} = \frac{\partial^2 f_i}{\partial y_j(i)\partial y_k(i)} = H_{jk}(i). \tag{18}$$

The local Newton update requires only the computation of local derivatives and the values of the neighbouring Lagrangian multipliers. The proposed local primal updates is given as:

$$\dot{y}_j(i) = \rho_d v_j(i) \tag{19}$$

$$\dot{v}_j(i) = -\gamma_d\left(\sum_{k=1}^{n}H_{jk}(i)v_k(i) + g_j(i)\right) \tag{20}$$

where ρ_d and γ_d are positive constants to be assigned. Since the local Hessians are all positive definite, it follows that the update for $v(i)$ has a unique equilibrium at $v(i) = -H(i)^{-1}g(i)$, which is the local expression of the Newton step for the minimization of the Lagrangian.

The network's primal dynamics can be written as follows:

$$\begin{aligned}\dot{\mathbf{y}} &= \rho_d\mathbf{v}\\ \dot{\mathbf{v}} &= -\gamma_d(\nabla^2_{yy}L(\mathbf{y},\boldsymbol{\lambda})\mathbf{v} + \nabla_y L(\mathbf{y},\boldsymbol{\lambda}))\end{aligned} \tag{21}$$

4. STABILITY ANALYSIS

In this section, we present an abridged version of the stability analysis. A detailed proof is outside the scope of the current study and will be presented in future work. The analysis of the proposed approach utilizes a time-scale separation. It is assumed that the update of the Lagrange multipliers $\boldsymbol{\lambda}$ is slow compared to the the primal local dynamics and the network dual dynamics.

Let us first summarize the overall closed-loop network dynamics:

$$\begin{aligned}\dot{\mathbf{y}} &= \rho_d\mathbf{v}\\ \dot{\mathbf{v}} &= -\gamma_d(\nabla^2_{yy}L(\mathbf{y},\boldsymbol{\lambda})\mathbf{v} + \nabla_y L(\mathbf{y},\boldsymbol{\lambda}))\\ \dot{\mathbf{z}} &= -\gamma(\mathcal{M}\mathbf{z} - \mathcal{M}\mathbf{y})\\ \dot{\boldsymbol{\delta}} &= -\gamma\left(\mathcal{M}\boldsymbol{\delta} + \nabla^2 f(\mathbf{y})\mathbf{z}\right)\\ \dot{\boldsymbol{\lambda}} &= \rho\boldsymbol{\delta}\end{aligned} \tag{22}$$

We first consider the gains as $\gamma_d = \frac{\gamma}{\varepsilon}$, $\rho_d = \frac{\rho}{\varepsilon}$ and $\gamma = \frac{\gamma}{\varepsilon}$. We then define the time-scale transformation $d\tau = \varepsilon dt$.

Let $\boldsymbol{y}(\boldsymbol{\lambda})$, $\boldsymbol{z}(\boldsymbol{\lambda})$ and $\boldsymbol{\delta}(\boldsymbol{\lambda})$ denote the equilibrium of the system for a given fixed $\boldsymbol{\lambda}$. As discussed above, the vector value function $\boldsymbol{y}(\boldsymbol{\lambda})$ is the unique solution of the system of equations:

$$\nabla_y L(\boldsymbol{y}(\boldsymbol{\lambda}), \boldsymbol{\lambda})) = 0. \tag{23}$$

The vectors $\boldsymbol{z}(\boldsymbol{\lambda})$ and $\boldsymbol{\delta}(\boldsymbol{\lambda})$ are the solutions of the following system of equations:

$$\mathcal{M}\boldsymbol{z}(\boldsymbol{\lambda}) = -\mathcal{M}\boldsymbol{y}(\boldsymbol{\lambda}), \tag{24}$$

$$\mathcal{M}(\nabla^2 f(\mathbf{y}(\boldsymbol{\lambda}))^{-1}\mathcal{M}\boldsymbol{\delta}(\boldsymbol{\lambda}) = \mathcal{M}\boldsymbol{z}(\boldsymbol{\lambda}). \tag{25}$$

Next we consider the error variables, $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{y}(\boldsymbol{\lambda})$, $\tilde{\boldsymbol{z}} - \boldsymbol{z}(\boldsymbol{\lambda})$ and $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda} - \boldsymbol{\delta}(\boldsymbol{\lambda})$. We can now write the error dynamics for the network in the new time-scale τ as follows:

$$\begin{aligned}
\frac{d\tilde{\mathbf{y}}}{d\tau} =& \rho_d \mathbf{v} - \frac{\partial \boldsymbol{y}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\frac{d\boldsymbol{\lambda}}{d\tau} \\
\frac{d\mathbf{v}}{d\tau} =& -\gamma_d(\nabla^2_{yy}L(\mathbf{y},\boldsymbol{\lambda})\mathbf{v} + \nabla_y L(\mathbf{y},\boldsymbol{\lambda})) \\
\frac{d\tilde{\mathbf{z}}}{d\tau} =& -\gamma(\mathcal{M}\mathbf{z} - \mathcal{M}\mathbf{y}) - \frac{\partial \boldsymbol{z}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\frac{d\boldsymbol{\lambda}}{d\tau} \\
\frac{d\tilde{\boldsymbol{\delta}}}{d\tau} =& -\gamma\Big(\mathcal{M}\boldsymbol{\delta} + \nabla^2 f(\mathbf{y})\mathbf{z}\Big) - \frac{\partial \boldsymbol{\delta}(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\frac{d\boldsymbol{\lambda}}{d\tau} \\
\frac{d\boldsymbol{\lambda}}{d\tau} =& \varepsilon\rho\boldsymbol{\delta}(\boldsymbol{\lambda}) + \varepsilon\rho\tilde{\boldsymbol{\delta}}.
\end{aligned} \tag{26}$$

The following theorem summarizes that the stability property of (26).

Theorem 1. Consider the Newton consensus algorithm expressed in error for (26). Assume that the communication network can be represented by undirected graph. Let $\gamma > 0$. Then there exists a constant ε^* such that for any $0 < \varepsilon < \varepsilon^*$, the origin of the system (26) have a globally asymptotically stable equilibrium at the origin.

Proof. We consider a singular perturbation analysis approach. We first define the boundary layer dynamics by setting $\varepsilon = 0$ in (**??**). This yields:

$$\begin{aligned}
\frac{d\tilde{\mathbf{y}}}{d\tau} =& \rho_d \mathbf{v} \\
\frac{d\mathbf{v}}{d\tau} =& -\gamma_d(\nabla^2_{yy}L(\mathbf{y},\boldsymbol{\lambda})\mathbf{v} + \nabla_y L(\mathbf{y},\boldsymbol{\lambda})) \\
\frac{d\tilde{\mathbf{z}}}{d\tau} =& -\gamma(\mathcal{M}\mathbf{z} - \mathcal{M}\mathbf{y}) \\
\frac{d\tilde{\boldsymbol{\delta}}}{d\tau} =& -\gamma\Big(\mathcal{M}\boldsymbol{\delta} + \nabla^2 f(\mathbf{y})\mathbf{z}\Big)
\end{aligned} \tag{27}$$

We first consider the first two equations that describes the primal dynamics of the network. The unique equilibrium of the systems occurs at $\boldsymbol{v} = 0$ and $\tilde{\boldsymbol{y}} = 0$. To assess the stability of the equilibrium, we pose the Lyapunov function candidate:

$$V = \frac{\rho_d}{2\gamma_d}\|\mathbf{v}\|^2 + L(\mathbf{y},\boldsymbol{\lambda}) - L(\mathbf{y}(\boldsymbol{\lambda}),\boldsymbol{\lambda}). \tag{28}$$

Its derivative is given by:

$$\dot{V} = -\rho_d \mathbf{v}^T \nabla^2_{yy} L(\mathbf{y},\boldsymbol{\lambda})\mathbf{v}.$$

Since the Hessian is positive definite for all $\tilde{y}$, it follows that $\dot{V} \le 0$ is negative semi-definite. The manifold $\Big\{(\mathbf{v},\, \tilde{y}) \,\Big|\, \dot{V} = 0\Big\}$ contains only one complete trajectory of the system which is the trivial point $\mathbf{v}^* = 0$, $\tilde{y} = 0$. By LaSalle's invariance principle, it follows that the equilibrium $(0,0)$ is a globally asymptotically stable equilibrium of the primal dynamics of the network.

The last two equations of (27) describe the consensus dynamics.Since the Hessian of the $f(\boldsymbol{y})$ is a constant, one can write the consensus dynamics (27) as follows:

$$\begin{aligned}
\frac{d\tilde{\mathbf{z}}}{d\tau} =& -\gamma(\mathcal{M}\tilde{\mathbf{z}} - \mathcal{M}\tilde{\mathbf{y}}) \\
\frac{d\tilde{\boldsymbol{\delta}}}{d\tau} =& -\gamma\Big(\mathcal{M}\tilde{\boldsymbol{\delta}} + \nabla^2 f(\mathbf{y}(\boldsymbol{\lambda}))\tilde{\mathbf{z}}\Big).
\end{aligned} \tag{29}$$

To perform the stability analysis, we consider the following coordinate transformations:

$$\tilde{\boldsymbol{z}} = ([\boldsymbol{r},\, \boldsymbol{R}] \otimes I_n)\boldsymbol{W},\; \tilde{\boldsymbol{y}} = ([\boldsymbol{r},\, \boldsymbol{R}] \otimes I_n)\boldsymbol{Y}.$$

and:

$$\tilde{\boldsymbol{\delta}} = ([\boldsymbol{r},\, \boldsymbol{R}] \otimes I_n)\boldsymbol{p}.$$

The dynamics are re-expressed in the new coordinates. The dynamics of $\boldsymbol{W}$ are given by:

$$\dot{\boldsymbol{W}} = -\gamma \begin{bmatrix} \boldsymbol{r}^T\mathcal{L}\boldsymbol{r} \otimes I_n & \boldsymbol{r}^T\mathcal{L}\boldsymbol{R} \otimes I_n \\ \boldsymbol{R}^T\mathcal{L}\boldsymbol{r} \otimes I_n & \boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_1 - \boldsymbol{Y}_1 \\ \boldsymbol{W}_2 - \boldsymbol{Y}_2 \end{bmatrix}$$

where $\boldsymbol{W} = \begin{bmatrix}\boldsymbol{W}_1^T & \boldsymbol{W}_2^T\end{bmatrix}$ and $\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y}_1^T & \boldsymbol{Y}_2^T\end{bmatrix}$. By the properties of $\boldsymbol{r}$ and $\boldsymbol{R}$, we obtain:

$$\begin{aligned}
\dot{\boldsymbol{W}}_1 &= 0 \\
\dot{\boldsymbol{W}}_2 &= -\gamma\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n(\boldsymbol{W}_2 - \boldsymbol{Y}_2).
\end{aligned}$$

Similarly, the dynamics of $\boldsymbol{p}$ is given by:

$$\begin{aligned}
\dot{\boldsymbol{p}}_1 &= -\gamma(\boldsymbol{r}^T \otimes I_n)\nabla^2 f(\boldsymbol{y}(\boldsymbol{\lambda}))([\boldsymbol{r}\;\; \boldsymbol{R}] \otimes I_n)\boldsymbol{W} \\
\dot{\boldsymbol{p}}_2 &= -\gamma(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{p}_2 \\
&\quad - \gamma(\boldsymbol{R}^T \otimes I_n)\nabla^2 f(\boldsymbol{y}(\boldsymbol{\lambda}))([\boldsymbol{r}\;\; \boldsymbol{R}] \otimes I_n)\boldsymbol{W}.
\end{aligned}$$

The Hessian matrix $\nabla^2 f(\boldsymbol{y}(\boldsymbol{\lambda}))$ can be written with respect to $\boldsymbol{Y}$ as:

$$\frac{\partial^2 f}{\partial \boldsymbol{Y} \partial \boldsymbol{Y}^T} = \left(\begin{bmatrix}\boldsymbol{r}^T \\ \boldsymbol{R}^T\end{bmatrix} \otimes I_n\right)\nabla^2 f(\boldsymbol{y}(\boldsymbol{\lambda}))([\boldsymbol{r}\;\; \boldsymbol{R}] \otimes I_n).$$

If one assumes that $W_1(0) = 0$, the dynamics of $\boldsymbol{p}$ can be written as:

$$\begin{aligned}
\dot{\boldsymbol{p}}_1 &= -\gamma \frac{\partial^2 f}{\partial \boldsymbol{Y}_1 \partial \boldsymbol{Y}_2^T}\boldsymbol{W}_2 \\
\dot{\boldsymbol{p}}_2 &= -\gamma(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{p}_2 - \gamma\frac{\partial^2 f}{\partial \boldsymbol{Y}_2 \partial \boldsymbol{Y}_2^T}\boldsymbol{W}_2.
\end{aligned}$$

Since the matrix $\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n$ is positive definite, then there exist constants $0 < \lambda_1 < \lambda_2$ such that:

$$\lambda_1 \left|\tilde{\boldsymbol{p}}_2\right|^2 \le \tilde{\boldsymbol{p}}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\tilde{\boldsymbol{p}}_2 \le \lambda_2 \left|\tilde{\boldsymbol{p}}_2\right|^2.$$

We can consider the Lyapunov function candidate: $V_1 = \frac{1}{2}\boldsymbol{W}_2^T\boldsymbol{W}_2 + \frac{\beta}{2}\boldsymbol{p}_2^T\boldsymbol{p}_2$ where β is a positive constant. This yields:

$$\begin{aligned}
\dot{V}_1 =& -\gamma\boldsymbol{W}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)(\boldsymbol{W}_2 - \boldsymbol{Y}_2) \\
& - \beta\gamma\boldsymbol{p}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{p}_2 - \beta\gamma\boldsymbol{p}_2^T\frac{\partial^2 f}{\partial \boldsymbol{Y}_2 \partial \boldsymbol{Y}_2^T}\boldsymbol{W}_2.
\end{aligned}$$

It is assumed that $\left|\frac{\partial^2 f}{\partial \mathbf{Y}_2 \partial \boldsymbol{Y}_2^T}\right| \le \bar{M}$. An upper bound for the right hand side can be computed using Young's inequality:

$$\begin{aligned}
\dot{V}_1 \le& -\gamma\boldsymbol{W}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{W}_2 - \beta\gamma\boldsymbol{p}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{p}_2 \\
& + \frac{\beta\gamma}{2k_1}\boldsymbol{p}_2^T\frac{\partial^2 f}{\partial \boldsymbol{Y}_2 \partial \boldsymbol{Y}_2^T}\boldsymbol{p}_2 + \frac{k_1\beta\gamma}{2}\boldsymbol{W}_2^T\frac{\partial^2 f}{\partial \boldsymbol{Y}_2 \partial \boldsymbol{Y}_2^T}\boldsymbol{p}_2\boldsymbol{W}_2 \\
& + \frac{\gamma}{2k_2}\boldsymbol{W}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{W}_2 \\
& + \frac{k_2\gamma}{2}\boldsymbol{Y}_2^T(\boldsymbol{R}^T\mathcal{L}\boldsymbol{R} \otimes I_n)\boldsymbol{Y}_2
\end{aligned}$$

This yields:

$$\begin{aligned}
\dot{V}_1 \le& -\left(\gamma\lambda_1 - \frac{k_1\beta\gamma\bar{M}}{2} - \frac{k_2\gamma\lambda_2}{2}\right)\boldsymbol{W}_2^T\boldsymbol{W}_2 \\
& - \left(\beta\gamma\lambda_1 - \frac{\beta\gamma\bar{M}}{2k_1}\right)\boldsymbol{p}_2^T\boldsymbol{p}_2 + \frac{k_2\gamma\lambda_2}{2}\boldsymbol{Y}_2^T\boldsymbol{Y}_2.
\end{aligned}$$

We set $k_1 = \frac{\bar{M}}{\lambda_1}$, $k_2 = \frac{\lambda_2}{\lambda_1}$ and $\beta = \frac{\lambda_1^2}{2M^2}$ and obtain:

$$\dot{V}_1 \le -\frac{\gamma\lambda_1}{4}\boldsymbol{W}_2^T\boldsymbol{W}_2 - \beta\frac{\gamma\lambda_1}{2}\tilde{\boldsymbol{p}}_2^T\tilde{\boldsymbol{p}}_2 + \frac{\gamma\lambda_2^2}{2\lambda_1}\boldsymbol{Y}_2^T\boldsymbol{Y}_2.$$

Therefore, the system is input-to-state stable (ISS) with $\boldsymbol{Y}_2$ as an input. From the previous analysis, the primal dynamics has an asymptotically stable equilibrium at $\tilde{\boldsymbol{y}} =$ 0m $\boldsymbol{v} = 0$. As a result, we see that system (27) is the interconnection of an asymptotically stable system and an ISS system. It follows from (Khalil (1992), Theorem 11.3) that the origin is a globally asymptotically stable equilibrium of (27).

Next we consider the reduced order system. We let $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\lambda})$ and write:

$$\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}t} = \varepsilon\rho\boldsymbol{\delta}(\boldsymbol{\lambda}). \tag{30}$$

Since $\boldsymbol{\delta}(\boldsymbol{\lambda})$ yields a Newton step for the dual problem, it follows that the reduced order system has a globally asymptotically stable equilibrium at the optimal value for Lagrange multiplier $\boldsymbol{\lambda}^*$.

Thus, we can apply a standard singular perturbation argument (Teel et al. (2003)) to conclude that there exists and ϵ^* such that, for all $0 < \epsilon < \epsilon^*$, the distributed network dynamics (22) has an asymptotically stable at the optimum solution of problem (1). This completes the proof.

5. DISTRIBUTED NEWTON SEEKING

In this section, we propose a Newton seeking approach to provide a model-free of the Newton consensus approach presented in the last sections. Locally, each agent implements the following primal optimization algorithm:

$$\dot{y}_j(i) = \rho_d v_j(i) \tag{31}$$

$$\dot{v}_j(i) = -\gamma_d\left(\sum_{k=1}^{n}\hat{H}_{jk}(i)v_k(i) + \hat{g}_j(i)\right) \tag{32}$$

where $\hat{H}_{jk}(i)$ is the estimation of $H_{jk}(i)$ (the jk element of the Hessian of f_i for agent i) and where

$$\hat{g}_j(i) = \hat{\xi}_j(i) + (\mathcal{L}\lambda_j)_i \tag{33}$$

is the local estimate of the gradient of the Lagrangian $L(\boldsymbol{y},\boldsymbol{\lambda})$ with respect to $y_j(i)$. Since the communication network Laplacian is assumed to be known, the second term on the right hand side of (33) is known. The first term, $\hat{\xi}_j(i)$, is a local estimate of the gradient of the local cost f_i for agent i with respect to $y_j(i)$.

For the purpose of this study, we consider a perturbation based Newton-seeking approach (Ghaffari et al., 2012). To do so, each agent must assign n frequencies, $\left[\omega_1^i\ \omega_2^i\ \dots\ \omega_n^i\right]$, such that:

i) $\omega_j^i \neq \omega_k^i$ for any $j \neq k$,
ii) ω_j^i/ω_k^i are rationals for any $j \neq k$,
iii) $\omega_j^i + \omega_k^i \neq \omega_\ell^i$ for all j, k, ℓ.

We let

$$s^i(t) = \left[a_1\sin(\omega_1^i t), \dots, a_n\sin(\omega_n^i t)\right]^T,$$

$$r^i(t) = \left[\tfrac{2}{a_1}\sin(\omega_1^i t), \dots, \tfrac{2}{a_1}\sin(\omega_n^i t)\right]^T,$$

and $\bar{y}(i) = [y_1(i), \dots, y_n(i)]^T$. We also define:

$$\Theta_{jj}^i(t) = \frac{16}{a_j^2}\left(\sin(\omega_j^i t)^2 - \frac{1}{2}\right), \tag{34}$$

$$\Theta_{jk}^i(t) = \frac{4}{a_j a_k}\sin(\omega_j^i t)\sin(\omega_k^i t) \tag{35}$$

where $j \neq k$.

Following Ghaffari et al. (2012), we consider the following ESC based approach for the primal problem:

$$\dot{y}_j(i) = \rho_d v_j(i) \tag{36}$$

$$\dot{v}_j(i) = -\gamma_d\left(\sum_{k=1}^{n}\hat{H}_{jk}(i)v_k(i) + \hat{g}_j(i)\right) \tag{37}$$

$$\dot{\hat{\xi}}_j(i) = -\omega_d\left(\hat{\xi}_j(i) - r_j^i(t)f_i(\bar{y}(i) + s^i(t))\right) \tag{38}$$

$$\dot{\hat{H}}_{jk}(i) = -\omega_d\left(\hat{H}_{jk}(i) - \Theta_{jk}^i(t)f_i(\bar{y}(i) + s^i(t))\right) \tag{39}$$

for $i = 1, \dots, m$, $j, k = 1, \dots, n$.

The dual problem is considered locally by implementing the following algorithm:

$$\dot{z}_j(i) = -\gamma\left(\sum_{k=1}^{m}\mathcal{L}_{ik}z_j(k) - \sum_{k=1}^{m}\mathcal{L}_{ik}y_j(k)\right) \tag{40}$$

$$\dot{\delta}_j(i) = -\gamma\left(\sum_{k=1}^{m}\mathcal{L}_{ik}\delta_j(k) + \sum_{k=1}^{n}\hat{H}_{jk}(i)z_k(i)\right) \tag{41}$$

$$\dot{\lambda}_j(i) = \rho\delta_j(i) \tag{42}$$

for $i = 1, \dots, m$, $j = 1, \dots, n$.

One can then generate an averaged realization of this system by integrating the right hand sides over the fundamental period of oscillation given by:

$$T(i) = 2\pi \times \mathrm{LCM}\left(\frac{1}{\omega_j^i}\right)$$

where LCM is the least common multiple of $\frac{1}{\omega_j^i}$.

The resulting averaged system recovers the dynamics of the Newton consensus algorithm presented in the previous sections up to averaged dynamics of the additional filters (38)-(39). The stability analysis for the Newton seeking is omitted. However it can be summarized as follows.

By the analysis in the previous section, it follows that the averaged system achieves Newton consensus and each agent have an asymptotically stable equilibrium at the minimizer of Problem (1). We can then use a standard averaging analysis (as in (Khalil, 1992)) to show that the Newton seeking system has a practically asymptotically stable equilibrium at the optimum and achieves a practical consensus at that optimum.

6. SIMULATION STUDY

To illustrate the ideas of the manuscript and the design of the method, we consider a simple 3 agent problem using the model based Newton consensus approach. The communication graph's Laplacian matrix is given by:

$$\mathcal{L} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

The local costs are given by:

$$\begin{aligned} f_1 &= y_1(1)^2 + (y_2(1) - 2)^2 \\ f_2 &= (y_1(2) - 1)^2 + y_2(2)^2 \\ f_3 &= 4 + (y_1(3) - 2)^2 + 2y_2(2)^2. \end{aligned}$$

The initial conditions are set to $\lambda_1(0) = \lambda_2(0) = [0\ 0\ 0]$ with all other variables set to 0 initially. The primal

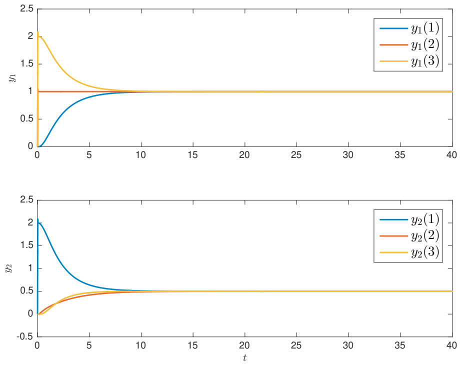

Fig. 1. Newton consensus results for Example 1 with tuning gains $\rho = 0.4$, $\gamma = 100$. The upper plot shows the consensus on $x_1^* = 1$ while the bottom shows consensus to $x_2^* = 0.5$.

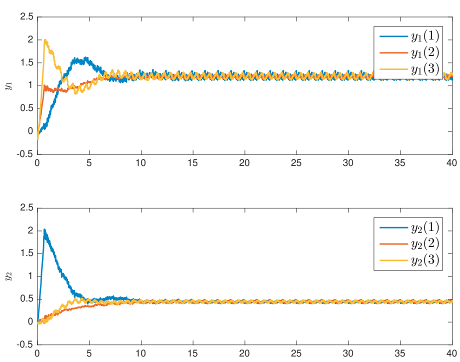

Fig. 2. Distributed Newton seeking results for Example 1 with tuning gains $\rho = 0.4$, $\gamma = 4$. The upper plot shows the consensus on $x_1^* = 1$ while the bottom shows consensus to $x_2^* = 0.5$.

dynamics gains are set to $\rho_d = \gamma_d = 100$. The dual gains are set to $\rho = 0.4$, $\gamma = 100$. The resulting consensus performance in given in Figure 1. The results show that the system correctly identifies the correct optimum $x^* = [1, 0.5]^T$. The consensus of y_1 to 1 and y_2 to 0.5 is achieved. The Newton consensus shows clearly that all agents converge to the optimum at exactly the same rate dictated by the gain ρ.

Next we apply the distributed Newton seeking approach. The gains used are the same as above. Two frequencies are required for each agent. They are as follows: $\omega_1^1 = 100$, $\omega_2^1 = 300$, $\omega_1^2 = 125$, $\omega_2^2 = 375$, $\omega_1^3 = 145$, $\omega_2^1 = 415$. The amplitudes are set to $A_1 = A_2 = A_3 = 0.5$. The low pass filter gain is set to $\omega_d = 200$. The results are shown in Figure 2. The Newton seeking achieves performance similar to the model-based technique in Figure 1.

7. CONCLUSION

This manuscript proposes a distributed Newton seeking for the solution of distributed optimization problems with locally measured but unknown cost functions. The approach implements a Newton consensus approach that can be implemented in a completely decentralized fashion. No exchange of derivative information between agents is required, and, no explicit inversion of the Hessian information is needed. The local gradients and Hessians are estimated using a Newton seeking control technique. Under the conditions considered in this study, the Newton seeking achieves a semi-global practical asymptotic stability of the optimum of the centralized optimization problem.

REFERENCES

Bertsekas, D.P. and Tsitsiklis, J.N. (1989). *Parallel and distributed computation*. Prentice Hall Inc., Old Tappan, NJ.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1–122.

Ghaffari, A., Krstić, M., and NešIć, D. (2012). Multivariable newton-based extremum seeking. *Automatica*, 48(8), 1759–1767.

Guay, M., Vandermeulen, I., et al. (2018). Distributed extremum-seeking control over networks of dynamically coupled unstable dynamic agents. *Automatica*, 93, 498–509.

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

Khalil, H. (1992). *Nonlinear Systems*. MacMillan Publishing Company, New York.

Michalowsky, S., Gharesifard, B., and Ebenbauer, C. (2018). On the lie bracket approximation approach to distributed optimization: Extensions and limitations. In *2018 European Control Conference (ECC)*, 119–124. IEEE.

Nedić, A. and Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1), 48–61.

Schizas, I., Ribeiro, A., and Giannakis, G. (2008). Consensus in ad hoc WSNs with noisy links ;part i: Distributed estimation of deterministic signals. *IEEE Trans. on Signal Processing,=*, 56(1), 350–364. doi: 10.1109/TSP.2007.906734.

Teel, A.R., Moreau, L., and Nesic, D. (2003). A unified framework for input-to-state stability in systems with two time scales. *IEEE Transactions on Automatic Control*, 48(9), 1526–1544. doi:10.1109/TAC.2003.816966.

Tutunov, R., Ammar, H.B., and Jadbabaie, A. (2019). Distributed newton method for large-scale consensus optimization. *IEEE Transactions on Automatic Control*.

Wei, E., Ozdaglar, A., and Jadbabaie, A. (2013). A distributed newton method for network utility maximization – part ii: Convergence. *IEEE Transactions on Automatic Control*, 58(9), 2176–2188.

Xu, J. and Soh, Y.C. (2013). Distributed extremum seeking control of networked large-scale systems under constraints. In *52nd IEEE Conf. on Decision and Control*, 2187–2192. doi:10.1109/CDC.2013.6760206.

Zanella, F., Varagnolo, D., Cenedese, A., Pillonetto, G., and Schenato, L. (2011). Newton-raphson consensus for distributed convex optimization. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 5917–5922. IEEE.