Lyapunov Analysis of a Distributed Optimization Scheme

Karla Kvaternik, Lacra Pavel

Lyapunov Analysis of a Distributed Optimization Scheme

Karla Kvaternik and Lacra Pavel

Abstract— **We analyze the convergence of the distributed multi-agent optimization scheme originally proposed in [1]. In this scheme, a number of agents cooperate to estimate the minimum of the sum of their locally-known cost functions. We consider a special case for which the collective cost function is strongly convex and where the agent communication graph is fixed. Whereas the analysis in [1] focuses on the suboptimality of the Cesàro averages of the agents' sequences, we establish explicit ultimate bounds on the agents' estimation errors themselves. We demonstrate that the collective optimum is globally practically asymptotically stable for this algorithm.**

I. INTRODUCTION

The development of distributed optimization schemes is an important and exciting research undertaking. Distributed optimization methods have many potential applications in the design of decentralized control protocols for systems that can be abstracted as a set of myopic agents which are required to accomplish a collective task while respecting communication limitations. Some examples of problems that can be addressed within a distributed optimization framework include congestion control and load balancing in IP networks [2], [3], power allocation in optical and wireless networks, [4], [5], voltage and frequency control in microgrids [6], and optimal sensor fusion in sensor networks [7].

In this paper we study a special case of the distributed optimization scheme proposed in the innovative work of Nedić and Ozdaglar in [1]. Therein, the authors study the convergence rate of a discrete-time algorithm that combines a (sub)gradient descent with a distributed averaging scheme commonly known as the *consensus algorithm* [8]. The algorithm proposed in [1] is executed by m agents, who try to agree on the minimizer of the sum of their individual costs. Each agent has knowledge only of his own cost function and exchanges his estimate of the collective minimizer with each of his neighbours over a time-varying communication graph.

A notable feature of this algorithm is that it can distributively solve resource allocation problems with non-separable (i.e. coupled) costs. The method thereby provides a cooperation-based alternative to game-theoretic approaches, which are known to suffer from the so-called "price of anarchy" in the absence of (often centralized) *mechanism designs*.

The convergence analysis given in [1] focuses on the *Cesàro averages* (i.e., time averages) of each agent's sequence of estimates. The main result therein provides a bound on the suboptimality of these Cesàro sequences at each iteration. Such sequences can be maintained locally by each agent without significant additional computational or memory requirements.

Unfortunately, in some control applications it may not be possible to make use of the Cesàro sequences. Since the convergence of a Cesàro sequence does not imply convergence of the original sequence, an important question to ask is whether there exists a neighbourhood of the collective minimizer to which the actual agent estimates themselves converge. If so, can this neighbourhood be characterized in terms of the problem parameters?

In this paper we provide an answer in the affirmative, under a slightly stronger set of assumptions than those made in [1]. Namely, we assume that the collective cost function is differentiable and strongly convex. On the other hand, we weaken the assumption that the subgradients of the agents' cost functions be bounded, requiring only that they be Lipschitz continuous. Furthermore, in this initial study, we consider only a fixed communication topology. Using Lyapunov techniques, we demonstrate that these assumptions suffice to guarantee the *practical asymptotic stability* of the collective minimizer. Additionally, our analysis is novel in that we view the evolution of the mean and deviation variables associated with the agent's estimates, as a feedback interconnection of two nonlinear subsystems. This approach allows us to work directly with the update equations, instead of working with algorithm solutions, as the authors in [1] do. To the best of our knowledge, this is the first completely Lyapunov-based analysis of the distributed optimization scheme proposed in [1].

In the next section we describe the problem setting, and state our objectives and assumptions. In Section III we provide the proof of our main result and discuss its consequences. We conclude our paper in Section IV.

A. General Notation

The set of non-negative real numbers is denoted $\mathbb{R}_+$. The Euclidean norm is denoted by $\|\cdot\|$, while $\|\cdot\|_1$ denotes the vector 1-norm. We often use $\mathbf{1} = [1,1,\ldots,1]^T \in \mathbb{R}^m$ and $\mathbf{0} = [0,0,\ldots,0]^T \in \mathbb{R}^m$. Given a matrix $P \in \mathbb{R}^{m\times n}$, $[P]_{i,j}$ is the entry of P in the ith row and jth column, while $[P]_i$ denotes the ith row of P. For a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$, $\nabla f(x)$ is its gradient at x. Occasionally we use x^+ to denote $x[k+1]$, and x for $x[k]$. For a function $F : \mathbb{R}^m \to \mathbb{R}$, we use the notation $\Delta F(x)$ to mean $F(x[k+1]) - F(x[k])$.

II. PROBLEM SETTING

In this paper we analyze a distributed optimization scheme involving m agents seeking to cooperatively solve the prob-

This work was supported by NSERC through the Vanier fund.

K. Kvaternik and L. Pavel are with the Edward S. Rogers Department of Electrical and Computer Engineering, University of Toronto {kvaternik,pavel}@utoronto.ca

lem

$$\min_{y\in\mathbb{R}} J(y), \tag{1}$$

where $J:\mathbb{R}\to\mathbb{R}$ is comprised of m additive components – i.e., $J(y)=J_1(y)+\ldots+J_m(y)$. We choose to work with a one-dimensional search space to ease our notation; our results readily generalize to multivariable optimization problems. We make the following assumption on this cost structure:

A2.1: (a): For each $i\in V$, $J_i(y)$ is differentiable with a gradient that satisfies the Lipschitz condition

$$\exists L_i>0 \quad \text{s.t.} \quad \|\nabla J_i(y_1)-\nabla J_i(y_2)\|\le L_i\|y_1-y_2\|, \tag{2}$$

For all $y_1,y_2\in\mathbb{R}$.

(b): The collective cost function $J(y)$ is strongly convex with constant $l>0$. $\diamondsuit$

In the sequel we make reference to the collective minimizer

$$y^* \triangleq \arg\min_{y\in\mathbb{R}} J(y),$$

and denote $J(y^*)$ by J^*. Strong convexity of J gives us the following two useful relationships [9]:

$$|\nabla J(y)|\ge l|y-y^*|, \quad \text{and} \tag{3}$$

$$|y-y^*|^2\le \tfrac{2}{l}\big(J(y)-J^*\big). \tag{4}$$

In this decentralized setting, agent i has knowledge only of J_i and may measure or otherwise evaluate ∇J_i at any point. However, he may communicate with some preordained subset of the other agents between each iteration.

The communication structure of the collective can be described by a weighted graph $G=(V,E(P))$, where $V=\{1,\ldots,m\}$ indexes the set of agents, $E(P)\subset V\times V$ is the set of communication links between them, with (j,i) belonging to E iff agent i receives information from agent j, and P is a matrix belonging to $\mathbb{R}_+^{m\times m}$, with $[P]_{i,j}>0$ iff $(j,i)\in E$.

The matrix P completely determines the communication structure of the multi-agent system. Although we assume that this structure is externally specified and not open to design, we require P to satisfy the following conditions:

A2.2: The matrix P is doubly stochastic, primitive, irreducible, and symmetric. $\diamondsuit$

The double stochasticity of P is equivalent to having $\mathbf{1}^TP=\mathbf{1}^T$ and that $P\mathbf{1}=\mathbf{1}$, and ensures that the spectral radius of P is $\rho(P)=1$ (c.f. Theorem 2.5.3 in [10]). Clearly, $\lambda_1(P)=1$ is an eigenvalue of P with eigenvector $\mathbf{1}$. The irreducibility of P is equivalent to the strong connectivity of G, and it guarantees that all eigenvalues of P with magnitude equal to one have multiplicity one (c.f. Theorem 2.1.4 (b) in [10]). Finally, the symmetry and primitivity of P ensure that there are no eigenvalues of P with magnitude one, other than $\lambda_1(P)$. Consequently, the second largest eigenvalue $\lambda_2(P)<1$.

It has been shown in [1] that problem (1) can be approximately solved by the Cesàro averages of the m sequences generated by the decentralized algorithm

$$x_i[k+1]=[P]_ix[k]-\alpha\nabla J_i(x_i[k]), \qquad \forall i\in V, \tag{5}$$

where $x_i[k]\in\mathbb{R}$ is agent i's estimate of y^* at the kth iteration, $x[k]\triangleq[x_1[k],\ldots,x_m[k]]^T$, and α is a constant step size. More precisely, for each $i\in V$, it has been shown that

$$J\Big(\frac{1}{k}\sum_{h=0}^{k-1}x_i[h]\Big)\le J^*+\frac{K_1}{k}+\alpha K_2, \tag{6}$$

for some constants K_1 and K_2 (c.f. Proposition 3 (b), [1]). Our aim is to demonstrate that under A2.1 and A2.2, each sequence $x_i[k]$ itself asymptotically converges to a neighbourhood of y^* that can be made arbitrarily small through the choice of the step size α.

III. MAIN RESULT

Before presenting our main result, we introduce some convenient notation and express the dynamics of algorithm (5) in terms of those of the mean of agents' estimates, and the deviation of these estimates from the mean.

Defining the vector field $d:\mathbb{R}^m\to\mathbb{R}^m$ by

$$d(x[k])\triangleq\begin{bmatrix}\nabla J_1(x_1[k])\\ \vdots\\ \nabla J_m(x_m[k])\end{bmatrix} \tag{7}$$

allows us to write (5) as

$$x[k+1]=Px[k]-\alpha d(x[k]). \tag{8}$$

We define two new variables,

$$y[k]\triangleq\tfrac{1}{m}\mathbf{1}^Tx[k], \tag{9}$$

and

$$z[k]\triangleq Mx[k], \tag{10}$$

where the matrix $M\in\mathbb{R}^{m\times m}$ is defined as

$$M\triangleq I-\phi, \qquad \text{and} \qquad \phi\triangleq\tfrac{1}{m}\mathbf{1}\mathbf{1}^T. \tag{11}$$

We note that $y[k]$ is the average of the agent estimates at iteration k, while the ith component of $z[k]$ represents the deviation of agent i's estimate of y^* from the average – i.e., $z[k]=x[k]-\mathbf{1}y[k]$. In [8], z is termed the "disagreement vector".

The double stochasticity of P (c.f. A2.2) implies that $\mathbf{1}^TP=\mathbf{1}^T$, and therefore we have that

$$\begin{aligned}y[k+1]&=\tfrac{1}{m}\mathbf{1}^TPx[k]-\tfrac{\alpha}{m}\mathbf{1}^Td(x[k])\\ &=y[k]-\tfrac{\alpha}{m}\mathbf{1}^Td(z[k]+\mathbf{1}y[k]).\end{aligned} \tag{12}$$

To write the difference equation that governs the evolution of the deviation vector, we first note that the relation $MP=PM$ follows readily from the definition of M and the double stochasticity of P. Then we observe that

$$\begin{aligned}z[k+1]&=M\big(Px[k]-\alpha d(x[k])\big)\\ &=PMx[k]-\alpha M\,d(x[k])\\ &=Pz[k]-\alpha M\,d(z[k]+\mathbf{1}y[k]).\end{aligned} \tag{13}$$

Expressions (12) and (13) form a system of coupled, nonlinear difference equations.

An essential idea in our proof is to show that (8) can be expressed as a perturbed version of the idealized system

$$\bar{x}[k+1] = P\bar{x}[k] - \alpha\phi\, d\big(\phi\bar{x}[k]\big). \tag{14}$$

which, unlike (8), has $\bar{x}[k] = \mathbf{1}y^*$ as its unique, asymptotically stable equilibrium. We demonstrate this assertion in the following Lemma.

Lemma 3.1: Suppose that assumptions A2.2 and A2.1 hold. Then, the point $\bar{x}[k] = \mathbf{1}y^*$ is the unique, asymptotically stable equilibrium of (14).

Proof: In the following we drop all time indices and use x^+ to denote $x[k+1]$.

(a): Uniqueness. To see that $\mathbf{1}y^*$ is an equilibrium for (14), we note that when $\bar{x} = \mathbf{1}y^*$, (14) becomes

$$\begin{aligned}\bar{x}^+ &= P\mathbf{1}y^* - \tfrac{\alpha}{m}\mathbf{1}\big(\mathbf{1}^T d(\tfrac{1}{m}\mathbf{1}\mathbf{1}^T\mathbf{1}y^*)\big)\\ &= \mathbf{1}y^* - \tfrac{\alpha}{m}\mathbf{1}\big(\nabla J(y^*)\big)\\ &= \bar{x},\end{aligned}$$

since the second term vanishes by the first order necessary condition for optimality.

To see that the equilibrium is unique, we define the matrix $Q \triangleq P - I$ and the subspace $\Gamma \triangleq \text{span}\{\mathbf{1}\}$. Suppose there exists another equilibrium point $p \in \mathbb{R}^m$, $p \neq \mathbf{1}y^*$. Then, p must satisfy $Qp = \alpha\phi d(\phi p)$. Either $p \in \Gamma$, or $p \notin \Gamma$. If $p \in \Gamma$, then $\exists c \in \mathbb{R}$ such that $p = \mathbf{1}c$. Since $Q = P - I$, we see that $Q\mathbf{1} = 0$, and therefore $Qp = 0$. Therefore, if $p \in \Gamma$ and is an equilibrium of (14), then it must be that $\mathbf{1}^T d(\mathbf{1}c) = \nabla J(c) = 0$, which, by the strong convexity of the J_i, $\forall i \in V$, is impossible unless $c \equiv y^*$. On the other hand, if $p \notin \Gamma$, then it must be that Qp is a non-zero vector in Γ. To see this, let $p_{av} := \frac{1}{m}\mathbf{1}^T p$, and write $\alpha\phi d(\phi p) = \frac{\alpha}{m}\mathbf{1}\big(\mathbf{1}^T d(\mathbf{1}p_{av})\big) := \mathbf{1}c_2 \in \Gamma$, for some scalar $c_2 \neq 0$. But since $\mathbf{1}^T Q = \mathbf{0}^T$, we see that $Qp = \mathbf{1}c_2$ implies that $0 = mc_2$, a contradiction.

(b): Asystability. To show the asystability of $\mathbf{1}y^*$ for (14), we examine the evolution of the mean and deviation of the vector $\bar{x}$, defined analogously to that of x in (9) and (10). We can show that these variables evolve according to the equations

$$\bar{y}^+ = \bar{y} - \tfrac{\alpha}{m}\mathbf{1}^T d(\mathbf{1}\bar{y}) \tag{15}$$

$$\bar{z}^+ = P\bar{z}, \tag{16}$$

which are decoupled, allowing us to study their stability properties separately.

(i): $\bar{y} \to y^*$. To show that $\bar{y} \to y^*$, it suffices to note that $\mathbf{1}^T d(\mathbf{1}\bar{y}) = \nabla J(y)$, and that by A2.1, ∇J is Lipschitz with the constant $L = \max_i L_i$. Then, we may apply a well-known result such as Theorem 1, §1.4 in [9], which states that $\bar{y}[k] \to y^*$ provided that $\frac{\alpha}{m} \in (0, \frac{L}{2})$.

(ii): $\bar{z} \to \mathbf{0}$. To show that $\bar{z} \to \mathbf{0}$, we define the Lyapunov function candidate $W(\bar{z}) \triangleq \bar{z}^T\bar{z}$, so that $W(\bar{z}^+) = \bar{z}^T(P^TP)\bar{z}$. Next, we note that $\mathbf{1}^T M = \mathbf{0}^T$, by the definition of M in (11). Since $\bar{z} \triangleq M\bar{x}$, it follows that $\mathbf{1}^T\bar{z}[k] \equiv 0$, $\forall k \in \mathbb{N}$; in other words, $\bar{z}$ is constrained to evolve on the orthogonal complement of $\Gamma = \text{span}\{\mathbf{1}\}$. As a result, $\bar{z}^T(P^TP)\bar{z} \leq \lambda_2(P)^2\|\bar{z}\|^2$, and by A2.2, $\lambda_2(P) < \lambda_1(P) = 1$. Using the notation $\Delta W(\bar{z}) \triangleq W(\bar{z}^+) - W(\bar{z})$ and defining the positive number $\mu \triangleq 1 - \lambda_2(P)^2$, we write

$$\begin{aligned}\Delta W(\bar{z}) &= \bar{z}^T P^T P\bar{z} - \bar{z}^T\bar{z}\\ &\leq -\mu\|\bar{z}\|^2.\end{aligned} \tag{17}$$

Then, since W is a positive definite, radially unbounded Lyapunov function with a negative definite difference ΔW, we conclude that $\mathbf{0}$ is an asymptotically stable equilibrium for (16) (c.f. Corollary 5.9.10 in [11], for example). ■

Remark 3.1: The convergence of the "sample variance" $z^T z$ in part *(b), (ii)* of Lemma 3.1 is studied also in [8] and [12], in the context of distributed averaging. ♢

Now that we have analysed the behaviour of the mean and deviation of the idealized algorithm, we show how to model (12) and (13) as perturbed versions thereof. In the next Lemma, we show that the perturbing terms in each case enter additively, and have favourable growth properties that ultimately enable our Lyapunov analysis.

Lemma 3.2: Suppose that part *(a)* of A2.1 holds, and let $L = \max_i L_i$. Then, system (12)-(13) can be expressed as

$$y^+ = y - \tfrac{\alpha}{m}\mathbf{1}^T d(\mathbf{1}y) - g(y,z) \tag{18}$$

$$z^+ = Pz - p(y,z) - \alpha C, \tag{19}$$

where $C = d(\mathbf{1}y^*)$,

$$|g(y,z)| \leq L\tfrac{\alpha}{\sqrt{m}}\|z\| \tag{20}$$

and

$$\|p(y,z)\| \leq \alpha L\|z\| + \alpha L\sqrt{m}|y - y^*|. \tag{21}$$

Proof: By adding and subtracting $\frac{\alpha}{m}\mathbf{1}^T d(\mathbf{1}y)$ to the right-hand side of (12), we obtain (18), with $g(y,z) = \frac{\alpha}{m}\mathbf{1}^T\big(d(z + \mathbf{1}y) - d(\mathbf{1}y)\big)$. Then we have

$$|g(y,z)| = \tfrac{\alpha}{m}\Big|\sum_{i=1}^m \nabla J_i(z_i + y) - \nabla J_i(y)\Big| \leq \tfrac{\alpha}{m}L\sum_{i=1}^m |z_i|, \tag{22}$$

and notice that (20) is obtained from the equivalence of norms – i.e., $\sum_{i=1}^m |z_i| = \|z\|_1 \leq \sqrt{m}\|z\|$.

To obtain (19), we first add and subtract the term $\alpha M\, d(\mathbf{1}y^*)$ from the right-hand side of (13), and let $p(y,z) = \alpha M\big(d(z + \mathbf{1}y) - d(\mathbf{1}y^*)\big)$ and $C = M\, d(\mathbf{1}y^*)$. By the first order necessary condition for optimality, $\mathbf{1}^T d(\mathbf{1}y^*) = 0$, and therefore $\phi d(\mathbf{1}y^*) = \mathbf{0}$. Since $M = I - \phi$, $C = d(\mathbf{1}y^*)$, as desired.

To bound $p(y,z)$, we first note that the spectral radius of M is one. Then,

$$\begin{aligned}\|p(y,z)\| &\leq \alpha\|M\|\|d(z + \mathbf{1}y) - d(\mathbf{1}y) + d(\mathbf{1}y) - d(\mathbf{1}y^*)\|\\ &\leq \alpha\|d(z + \mathbf{1}y) - d(\mathbf{1}y)\| + \alpha\|d(\mathbf{1}y) - d(\mathbf{1}y^*)\|\\ &\leq \alpha L\|z\| + \alpha\sqrt{m}L|y - y^*|,\end{aligned} \tag{23}$$

which was to be shown. ■

With this analysis of algorithm (8), we are ready to state our main result. In order to provide its proof, we need a technical tool, which we summarize in the following Lemma, whose proof can also be found in [9]:

Lemma 3.3: Given any differentiable function $g : \mathbb{R}^m \to \mathbb{R}$

and any $a,b \in \mathbb{R}^m$, we can write

$$g(a+b) = g(a) + \nabla g(a)^T b + \int_0^1 [\nabla g(a+\tau b) - \nabla g(a)]^T b \, d\tau. \tag{24}$$

Proof: Let $u(\tau) = a + \tau b$. Then by the chain rule,

$$\tfrac{d}{d\tau} g(a+\tau b) = \nabla_u g(u)^T \nabla_\tau u(\tau) = \nabla g(a+\tau b)^T b.$$

By Leibniz's rule, $\int_0^1 \frac{d}{d\tau} g(a+\tau b)\, d\tau = g(a+b) - g(a)$, and therefore

$$g(a+b) = g(a) + \int_0^1 \nabla g(a+\tau b)^T b \, d\tau. \tag{25}$$

The result then follows by adding and subtracting $\nabla g(a)^T b$ to the right-hand side of (25). ■

In the following theorem, we carry out a Lyapunov analysis for the system (18)-(19), relative to the point $[y, z^T]^T = [y^*, \mathbf{0}^T]^T$. We use the notation $\Delta V(y,z) = V(y[k+1], z[k+1]) - V(y[k], z[k])$.

Theorem 3.1: Suppose that A2.1 and A2.2 hold. Then, the Lyapunov function $V(y,z) = z^T z + \frac{2m}{l}(J(y) - J^*)$ decreases along the sequences generated by (18)-(19) as

$$\Delta V(y,z) \le -K_z \|z\|^2 - K_y |y-y^*|^2 + \alpha^2 K, \tag{26}$$

for some $K_z, K_y \in \mathbb{R}$, and $K > 0$. Moreover, K_z and K_y are positive whenever α satisfies $0 < \alpha < \min\{\bar{\alpha}_1, \bar{\alpha}_2\}$, where

$$\bar{\alpha}_1 = -\tfrac{m(4Ll^2 + 8(mL)^2 + L^4)}{4(Ll)^2(2m+L)} + \sqrt{\left(\tfrac{m(4Ll^2+8(mL)^2+L^4)}{4(Ll)^2(2m+L)}\right)^2 + \tfrac{m(1-\lambda_2(P)^2)}{2L^2(2m+L)}} \tag{27}$$

and

$$\bar{\alpha}_2 = \tfrac{1}{4}\left(\tfrac{L}{m} + 2\left(\tfrac{mL}{l}\right)^2 + 1\right)^{-1} \tag{28}$$

Proof: First we examine the changes in the value of J along the sequence generated by (18). To this end, we expand J about y, with a deviation $b = \frac{\alpha}{m}\mathbf{1}^T d(\mathbf{1}y) - g(y,z)$. According to Lemma 3.3, we obtain

$$\begin{aligned} J(y+b) &= J(y) + \nabla J(y) b + \int_0^1 \left[\nabla J(y+\tau b) - \nabla J(y)\right] b \, d\tau \\ &\le J(y) + \nabla J(y) b + \tfrac{L}{2}|b|^2, \end{aligned} \tag{29}$$

where the inequality is obtained from the Lipschitz continuity of ∇J. Let $\Delta J(y)$ denote the difference $J(y^+) - J(y)$. Then, substituting for b and noting that

$$|b|^2 = \left|\tfrac{\alpha}{m}\mathbf{1}^T d(\mathbf{1}y) - g(y,z)\right|^2 \le 2\left(\tfrac{\alpha}{m}\nabla J(y)\right)^2 + 2g(y,z)^2,$$

we write

$$\begin{aligned} \Delta J(y) &\le -\tfrac{\alpha}{m}\left(1 - L\tfrac{\alpha}{m}\right)|\nabla J(y)|^2 - \nabla J(y) g(y,z) + L|g(y,z)|^2 \\ &\le -\tfrac{\alpha l^2}{m}\left(1 - L\tfrac{\alpha}{m}\right)|y-y^*|^2 + L|y-y^*|\,|g(y,z)| \\ &\quad + L|g(y,z)|^2, \end{aligned}$$

where we obtain the last inequality by applying (3) to the first term and A2.1 (a) to the second term, with $L = \max_i L_i$.

Next, using (20), we obtain

$$\begin{aligned} \Delta J(y) &\le -\tfrac{\alpha l^2}{m}\left(1 - L\tfrac{\alpha}{m}\right)|y-y^*|^2 + \tfrac{L^2\alpha}{\sqrt{m}}|y-y^*|\,\|z\| + \tfrac{L^3\alpha^2}{m}\|z\|^2 \\ &\le -\tfrac{\alpha l^2}{m}\left(\tfrac{1}{2} - L\tfrac{\alpha}{m}\right)|y-y^*|^2 + \left(\tfrac{L^3\alpha^2}{m} + \tfrac{\alpha L^4}{2l^2}\right)\|z\|^2 \end{aligned} \tag{30}$$

where we have used the fact that $ab \le \frac{1}{2}a^2 + \frac{1}{2}b^2$ and combined terms to obtain the last inequality.

Next, we turn to the deviation subsystem, and examine the changes in the value of the function $W(z) \triangleq z^T z$ along the trajectories of (19). By A2.2,

$$\begin{aligned} W(z^+) &= (Pz - p(y,z) - \alpha C)^T (Pz - p(y,z) - \alpha C) \\ &\le \lambda_2(P)^2\|z\|^2 + \|p(y,z)\|^2 + \alpha^2\|C\|^2 \\ &\quad + 2\|z\|\|p(y,z)\| + 2\alpha\|z\|\|C\| + 2\alpha\|p(y,z)\|\|C\|, \end{aligned}$$

where we have used the sub-multiplicativity of the induced 2-norm of the matrix P, and the fact that $\|P\| \le 1$. We define the positive constant $\mu \triangleq 1 - \lambda_2(P)$. Then, using (21) and the fact that $\|p(y,z)\|^2 \le 2(\alpha L)^2\|z\|^2 + 2m(\alpha L)^2|y-y^*|^2$, we write

$$\begin{aligned} \Delta W(z) &\le -\mu\|z\|^2 + 2(\alpha L)^2\|z\|^2 + 2m(\alpha L)^2|y-y^*|^2 \\ &\quad + \alpha^2\|C\|^2 + 2\alpha L\|z\|^2 + 2\alpha L\sqrt{m}\|z\||y-y^*| \\ &\quad + 2\alpha(1+\alpha L)\|C\|\|z\| + 2\alpha^2 L\sqrt{m}\|C\||y-y^*|. \end{aligned} \tag{31}$$

Next, we apply the inequality $ab \le \varepsilon a^2 + \frac{1}{4\varepsilon}b^2$ to the three terms involving $\|z\|\,|y-y^*|$, $\|z\|$ and $|y-y^*|$, expressing them in terms of constants and terms in $\|z\|^2$ and $|y-y^*|^2$. Combining terms, we obtain

$$\begin{aligned} \Delta W(z) &\le -\left(\mu - 2\alpha L - 2(\alpha L)^2 - \tfrac{\alpha}{\varepsilon_1}\left(\tfrac{mL}{l}\right)^2 - \varepsilon_2\right)\|z\|^2 \\ &\quad + \left(2m(\alpha L)^2 + \varepsilon_1\tfrac{\alpha l^2}{m} + \alpha\varepsilon_3\tfrac{\alpha l^2}{m}\right)|y-y^*|^2 \\ &\quad + \left(\alpha^2\|C\|^2 + \tfrac{\alpha^2(1+\alpha L)^2\|C\|^2}{\varepsilon_2} + \tfrac{\alpha^2(Lm\|C\|)^2}{\varepsilon_3 l^2}\right) \end{aligned} \tag{32}$$

Finally, we combine equations (30) and (32) to form $\Delta V(y,z)$:

$$\begin{aligned} \Delta V(y,z) &= \Delta W(z) + \tfrac{2m}{l}\Delta J(y) \\ &\le -\Big(\mu - 2(\alpha L)^2 - 2\alpha L - \tfrac{\alpha}{\varepsilon_1}\left(\tfrac{mL}{l}\right)^2 \\ &\qquad - \varepsilon_2 - \tfrac{\alpha^2 L^3}{m} - \tfrac{\alpha L^4}{2l^2}\Big)\|z\|^2 \\ &\quad - 2\alpha l\Big(\tfrac{1}{2} - L\tfrac{\alpha}{m} - 2\alpha\left(\tfrac{mL}{l}\right)^2 - \varepsilon_1 - \alpha\varepsilon_3\Big)|y-y^*|^2 \\ &\quad + \alpha^2\Big(\|C\|^2 + \tfrac{(1+\alpha L)^2\|C\|^2}{\varepsilon_2} + \tfrac{(Lm\|C\|)^2}{\varepsilon_3 l^2}\Big). \end{aligned}$$

Choosing $\varepsilon_1 = \frac{1}{4}$, $\varepsilon_2 = \frac{1}{2}\mu$ and $\varepsilon_3 = 1$ gives us the required form (26), with

$$K = \|d(\mathbf{1}y^*)\|^2\left(1 + \tfrac{2(1+\alpha L)^2}{1-\lambda_2(P)^2} + \left(\tfrac{mL}{l}\right)^2\right), \tag{33}$$

$$\begin{aligned} K_z &= \tfrac{1}{2}(1 - \lambda_2(P)^2) - \alpha^2\left(2L^2 + \tfrac{L^3}{m}\right) \\ &\quad - \alpha\left(2L + 4\left(\tfrac{mL}{L}\right)^2 + \tfrac{L^4}{2l^2}\right), \end{aligned} \tag{34}$$

$$K_y = 2\alpha l\left(\tfrac{1}{4} - \alpha\left(\tfrac{L}{m} + 2\left(\tfrac{mL}{l}\right)^2 + 1\right)\right), \tag{35}$$

where we have also recalled the fact that $C = d(\mathbf{1}y^*)$ and $\mu = 1 - \lambda_2(P)^2$. From these expressions, it can easily be verified that K_y and K_z are rendered positive whenever

$\alpha \in (0, \min\{\bar{\alpha}_1, \bar{\alpha}_2\})$, with $\bar{\alpha}_1$ and $\bar{\alpha}_2$ as in (27) and (28). ■

Remark 3.2: Our Lyapunov function $V(y[k], z[k]) = z[k]^T z[k] + \frac{2m}{l}(J(y[k]) - J^*)$ is a joint measure of the sample variance of agents' estimates, and the suboptimality of the mean of their estimates at the kth iteration. Theorem 3.1 tells us that V decreases as (8) evolves until $K_z\|z\|^2 + K_y|y - y^*|^2 \leq \alpha^2 K$.

Since J is assumed to be strongly convex (c.f. A2.1), we may relate the error in the agent's estimates to V as

$$\begin{aligned}\|x - \mathbf{1}y^*\|^2 &= \|z + \mathbf{1}y - \mathbf{1}y^*\|^2 \\ &\leq 2\|z\|^2 + 2m|y - y^*|^2 \\ &\leq 2\|z\|^2 + 2\tfrac{2m}{l}(J(y) - J^*) \\ &= 2V(y, z),\end{aligned}$$

where we have obtained the last inequality by applying (4).

But to what value does the sequence $V(y[k], z[k])$ converge? The number $\bar{V} \triangleq \limsup_{k\to\infty} V(y[k], z[k])$ can be established as the maximum value that V attains on the set

$$Z = \{(y, z^T)^T \in \mathbb{R}^{m+1} \,\big|\, K_z\|z\|^2 + K_y|y - y^*|^2 \leq \alpha^2 K\}, \quad (36)$$

on which V ceases to decrease since $\Delta V(y, z)$ is no longer negative. Such a value is guaranteed to exist since Z is compact and V is continuous. Specifically, we find that

$$\lim_{k\to\infty} \|x[k] - \mathbf{1}y^*\|^2 \leq 2 \lim_{k\to\infty} V(y[k], z[k]) \leq 2\bar{V}, \quad (37)$$

where

$$\bar{V} = \max_{(y,z^T)^T \in Z} V(y, z) = \max_{(y,z^T)^T \in Z} z^T z + \tfrac{2m}{l} \max_{(y,z^T)^T \in Z} (J(y) - J^*).$$

The second term can be bounded by using the relationship $J(y_o + b) \leq J(y_o) + \nabla J(y_o)b + \frac{L}{2}|b|^2$ (c.f. (29) from Lemma 3.3), with $y_o = y^*$ and $b = y - y^*$. We thus obtain

$$\bar{V} \leq \max_{(y,z^T)^T \in Z} z^T z + \tfrac{mL}{l} \max_{(y,z^T)^T \in Z} |y - y^*|^2.$$

By the definition of Z in (36), we see that the first term above is maximized when $|y - y^*| = 0$, while the second term is maximized when $\|z\| = 0$. Therefore, an ultimate upper bound on the agent's actual estimate errors is given by

$$\lim_{k\to\infty} \|x[k] - \mathbf{1}y^*\|^2 \leq 2\alpha^2 \tfrac{K}{K_z} + 2\alpha^2 \tfrac{mL}{l} \tfrac{K}{K_y}, \quad (38)$$

where K_z, K_y and K are as in (33) to (35). The most important observation to be made concerning (38) is the fact that the estimation error can be made arbitrarily small through the parameter α. $\diamondsuit$

Remark 3.3: Let x_i^* denote the minimizer of J_i. It is interesting to note that if all the individual agent optima coincide, – i.e., $x_i^* = y^*$, $\forall i \in V$, then $K \equiv 0$ by the first-order necessary condition for optimality. Consequently, $\lim_{k\to\infty} \|x[k] - \mathbf{1}y^*\| = 0$, even for a fixed step size α. $\diamondsuit$

Remark 3.4: In order to prove that $(y^*, \mathbf{0}^T)^T$ is *practically asymptotically stable* for the system (18)-(19), we would need to demonstrate the existence of a compact, positively invariant set $P \subset \mathbb{R}^{m+1}$, containing $(y^*, \mathbf{0}^T)^T$ and having the property that all sequences initiated in $\mathbb{R}^{m+1} \backslash P$ enter P in finite time. It can be shown using techniques similar to those found in the proof of Theorem 5.14.2 in [11], that for any arbitrarily small $\delta > 0$, $P = \{(y, z^T)^T \in \mathbb{R}^{m+1} \,\big|\, V(y, z) \leq \bar{V} + \alpha^2 K + \delta\}$ is such a set. $\diamondsuit$

IV. CONCLUSIONS

In this paper we have studied the convergence properties of a distributed multi-agent optimization algorithm. The algorithm we considered involves a number of agents that communicate over a fixed information exchange graph and cooperate to estimate a collective cost minimizer. Using Lyapunov techniques we derived explicit ultimate bounds on the agents' estimation errors and showed that these errors diminish with a tunable algorithm parameter. Our future work will focus on allowing for a time-varying information exchange graph and asynchronous operation of the agents.

REFERENCES

[1] A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54:48–60, 2009.

[2] F.P. Kelly, A.K. Maulloo, and D.K.H. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operations Research Society*, 49:237–252, 1998.

[3] S. Kandula, D. Katabi, B. Davie, and A. Charny. Walking the tihgtrope: responsive yet stable traffic engineering. In *SIGCOMM'05, PA, USA*, 2005.

[4] L. Pavel. An extension of duality to a game-theoretic framework. *Automatica*, 43:226–237, 2007.

[5] X. Fan, T. Alpcan, M. Arcak, T.J. Wen, and T. Başar. A passivity based approach to game-theoretic cdma power control. *Automatica*, 42:1837–1847, 2006.

[6] A. Mehrizi-Sani and R. Iravani. Potential-function based control of a microgrid in islanded and grid-connected modes. *IEEE Transactions on Power Systems*, 25:1883–1891, 2010.

[7] S. Sundar Ram, V.V. Veeravalli, and A. Nedić. Distributed and recursive parameter estimation in parametrized linear state-space models. *IEEE Transactions on Automatic Control*, 55:488–492, 2010.

[8] R.O. Saber, J.A. Fax, and R.M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95:215–233, 2007.

[9] B.T. Polyak. *Introduction to optimization*. Optimization Software, Inc., 1987.

[10] A. Berman and R.J. Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, 1979.

[11] R.P. Agarwal. *Difference equations and inequalities - theory, methods and applications*. Marcel Dekker, 2000.

[12] A. Nedić, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, 54:2506–2517, 2009.