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

Distributed Subgradient Methods and Quantization Effects

Angelia Nedić
Department of Industrial and
Enterprise Systems Engineering
University of Illinois
Urbana-Champaign, IL 61801
Email: angelia@illinois.edu

Alex Olshevsky, Asuman Ozdaglar, and John N. Tsitsiklis
Department of Electrical Engineering and Computer Science
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02142
Email: {alex_o, asuman, jnt}@mit.edu

***Abstract*— We consider a convex unconstrained optimization problem that arises in a network of agents whose goal is to cooperatively optimize the sum of the individual agent objective functions through local computations and communications. For this problem, we use averaging algorithms to develop distributed subgradient methods that can operate over a time-varying topology. Our focus is on the convergence rate of these methods and the degradation in performance when only quantized information is available. Based on our recent results on the convergence time of distributed averaging algorithms, we derive improved upper bounds on the convergence rate of the unquantized subgradient method. We then propose a distributed subgradient method under the additional constraint that agents can only store and communicate quantized information, and we provide bounds on its convergence rate that highlight the dependence on the number of quantization levels.**

I. Introduction

There has been much interest in developing distributed methods for optimization in networked-systems consisting of multiple agents with local information structures. Such problems arise in a variety of environments including resource allocation among heterogeneous agents in large-scale networks, and information processing and estimation in sensor networks. Optimization algorithms deployed in such networks should be completely distributed, relying only on local observations and information, and robust against changes in network topology due to mobility or node failures.

Recent work [15] has proposed a subgradient method for optimizing the sum of convex objective functions corresponding to n agents connected over a time-varying topology (see also the short paper [14]). The goal of the agents is to cooperatively solve the unconstrained optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f_i(x) \\ \text{subject to} & x \in \mathbb{R}^m, \end{array} \tag{1}$$

where each $f_i : \mathbb{R}^m \to \mathbb{R}$ is a convex function, representing the local objective function of agent i, and known only to this agent. The decision vector x in problem (1) can be viewed as either a resource vector whose components correspond to resources allocated to each agent, or a global estimate vector to be optimized by the agents using local information. For example, such a problem arises in distributed sensor networks where the sensors are spatially distributed over a field to measure and estimate certain quantities. The objective function of sensor i has a form $f_i(x) = \mathrm{E}[F_i(R_i, x)]$, where R_i is some random process observed locally by agent i, the function $F_i(R_i, x)$ captures the quality of agent i estimates, and E denotes the expectation (see [21]).

Our proposed method builds on the work in [24], [25] (see also, [3]). It relies on every agent maintaining estimates of an optimal solution to problem (1), and communicating these estimates locally to its neighbors. The estimates are updated using a combination of a subgradient iteration[1] and an averaging algorithm. The subgradient step optimizes the local objective function while the averaging algorithm is used to obtain information about the objective functions of the other agents.

In this paper, we consider the distributed subgradient method discussed in [15], and provide improved convergence rate results. In particular, we use our recent results on the convergence time of averaging algorithms [13] and establish new upper bounds on the difference between the objective function value of the estimates of each agent and the optimal value of problem (1). These bounds have a polynomial dependence on the number of agents n (in contrast with the error bounds in [15], [14], which involve exponential dependence on n). Furthermore, we study a variation of the distributed subgradient method in

This research was partially supported by the National Science Foundation under grant ECCS-0701623, CAREER grants CMMI 07-42538 and DMI-0545910, and by DARPA ITMANET program.

[1]For subgradient methods see, for example, [19], [22], [20], [8], [1], [2].

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

which the agents have access to quantized information, and provide bounds on the convergence time that contain additional error terms due to quantization.

In addition to the papers cited above, our work is related to the literature on reaching consensus on a particular scalar value or on computing exact averages of the initial values of the agents, a subject motivated by natural models of cooperative behavior in networked-systems (see, e.g., [26], [9], [4], [16], [5], and [17], [18]). Closely related is also the work in [10] and [7], [6], which study the effects of quantization on the performance of averaging algorithms. Our work is also related to the *utility maximization* framework for resource allocation in networks (see [11], [12], [23]). In contrast to this literature, however, we allow the local objective functions to depend on the entire resource allocation vector.

The rest of this paper is organized as follows. In Section II, we describe the distributed subgradient method and present an improved convergence rate estimate using our recently established bounds on the convergence time of our averaging algorithms [13]. In Section III, we consider a version of the method under the additional constraint that the agents can only exchange quantized information. We provide convergence and rate of convergence results as a function of the number of quantization levels. Section IV contains our concluding remarks.

Notation and Basic Notions. We view all vectors as columns. We use e_i to denote the vector with ith entry equal to 1 and all other entries equal to 0. We use $\mathbf{1}$ to denote a vector with all entries equal to 1. For a matrix A, we use a_{ij} or $[A]_{ij}$ to denote the matrix entry in the ith row and jth column. We write $[A]_i$ and $[A]^j$ to denote respectively the ith row and the jth column of a matrix A. A vector a is said to be a *stochastic vector* when its components a_i are nonnegative and $\sum_i a_i = 1$. A square matrix A is said to be *stochastic* when each row of A is a stochastic vector, and it is said to be *doubly stochastic* when both A and its transpose A' are stochastic matrices.

For a convex function $F : \mathbb{R}^m \to \mathbb{R}$, we use the notion of a subgradient (see [2]): a vector $s_F(\bar{x}) \in \mathbb{R}^m$ *is a subgradient of a convex function* F *at* $\bar{x}$ if

$$F(\bar{x}) + s_F(\bar{x})'(x - \bar{x}) \le F(x) \qquad \text{for all } x.$$

We use the notation $f(x) = \sum_{j=1}^n f_i(x)$. We denote the optimal value of problem (1) by f^* and the set of optimal solutions by X^*.

II. Distributed Subgradient Method

We first introduce our distributed subgradient method for solving problem (1) and discuss the assumptions imposed on the information exchange among the agents. We consider a set $V = \{1, \ldots, n\}$ of agents. Each agent starts with an initial estimate $x_i(0) \in \mathbb{R}^m$ and updates its estimate at discrete times $t_k, k = 1, 2, \ldots$. We denote by $x_i(k)$ the vector estimate maintained by agent i at time t_k. When updating, an agent i combines its current estimate x_i with the estimates x_j received from its neighboring agents j. Specifically, agent i updates its estimates by setting

$$x_i(k+1) = \sum_{j=1}^n a_{ij}(k) x_j(k) - \alpha d_i(k), \qquad (2)$$

where the scalars $a_{i1}(k), \ldots, a_{in}(k)$ are nonnegative weights and the scalar $\alpha > 0$ is a stepsize. The vector $d_i(k)$ is a subgradient of the agent i cost function $f_i(x)$ at $x = x_i(k)$. We use the notation $A(k)$ to denote the *weight matrix* $[a_{ij}(k)]_{i,j=1,\ldots,n}$.

The evolution of the estimates $x_i(k)$ generated by Eq. (2) can be equivalently represented using transition matrices. In particular, we define a *transition matrix* $\Phi(k, s)$ for any s and k with $k \ge s$, as follows:

$$\Phi(k, s) = A(k)A(k-1)\cdots A(s+1)A(s).$$

Using these transition matrices, we relate the estimate $x_i(k+1)$ to the estimates $x_1(s), \ldots, x_n(s)$ for any $s \le k$. In particular, for the iterates generated by Eq. (2), we have for any i, and any s and k with $k \ge s$,

$$\begin{aligned} x_i(k+1) \quad &= \quad \sum_{j=1}^n [\Phi(k,s)]_{ij} x_j(s) \\ &- \quad \alpha \sum_{r=s}^{k-1} \sum_{j=1}^n [\Phi(k, r+1)]_{ij} d_j(r) \\ &- \quad \alpha\, d_i(k) \end{aligned} \qquad (3)$$

(for more details, see [15]). As seen from the preceding relation, to study the asymptotic behavior of the estimates $x_i(k)$, we need to understand the behavior of the transition matrices $\Phi(k, s)$. We do this under some assumptions on the agent interactions that translate into some properties of transition matrices.

Our first assumption imposes some conditions on the weights $a_{ij}(k)$ in Eq. (2).

Assumption 1: For all $k \ge 0$, the weight matrix $A(k)$ is doubly stochastic with positive diagonal. Additionally, there is a scalar $\eta > 0$ such that if $a_{ij}(k) > 0$, then $a_{ij}(k) \ge \eta$.

The doubly stochasticity assumption on the weight matrix will guarantee that the subgradient of the objective function f_i of every agent i will receive the same weight in the long run. The second part of the assumption states that each agent gives significant weight to its own values and to the values of its neighbors.

Remark 1: For example, we can ensure in a distributed manner that the weight matrix $A(k)$ satisfies Assumption 1 when the agent communications are bidirectional. In this case, we allow each agent i to have planned weights $\tilde{a}_{ij}(k), j = 1 \ldots, n$ that the agent communicates to its neighbors together with the estimate $x_i(k)$, where the matrix $\tilde{A}(k)$ of planned weights is a (row) stochastic matrix satisfying Assumption 1, except for doubly stochasticity. In particular, at time k, if agent j communicates with agent i, then agent i receives $x_j(k)$ and the planned weight $\tilde{a}_{ji}(k)$ from agent j. At the same time, agent j receives $x_i(k)$ and the planned weight $\tilde{a}_{ij}(k)$ from agent i. Then, the actual weights that an agent i uses are given by

$$a_{ij}(k) = \min\{\tilde{a}_{ij}(k), \, \tilde{a}_{ji}(k)\},$$

if i and j talk at time k, and $a_{ij}(k) = 0$ otherwise; while

$$a_{ii}(k) = 1 - \sum_{\{j \mid j \leftrightarrow i \text{ at time } k\}} a_{ij}(k),$$

where the summation is over all j communicating with i at time k. It can be seen that the weights $a_{ij}(k)$ satisfy Assumption 1. Metropolis weights [27] are another example of weights satisfying Assumption 1.

At each time k, the agents' connectivity can be represented by a directed graph $G(k) = (V, \mathcal{E}(A(k)))$, where $\mathcal{E}(A)$ is the set of directed edges (j, i), including self-edges (i, i), such that $a_{ij}(k) > 0$. Our next assumption ensures that the agents are connected frequently enough to persistently influence each other.

Assumption 2: There exists an integer $B \geq 1$ such that the directed graph

$$\Big(V, \mathcal{E}(A(kB)) \cup \cdots \cup \mathcal{E}(A((k+1)B - 1))\Big)$$

is strongly connected for all $k \geq 0$.

A. Preliminary Results

Here, we provide some results that we use later in our convergence analysis of method (2). These results hold under Assumptions 1 and 2.

Consider a related update rule of the form

$$z(k+1) = A(k)z(k), \tag{4}$$

where $z(0) \in \mathbb{R}^n$ is an initial vector[2]. Define

$$V(k) = \sum_{j=1}^{n} (z_j(k) - \bar{z}(k))^2 \qquad \text{for all } k \geq 0,$$

where $\bar{z}(k)$ is the average of the entries of the vector $z(k)$. Under the doubly stochasticity of $A(k)$, the initial average $\bar{z}(0)$ is preserved by the update rule (4), i.e., $\bar{z}(k) = \bar{z}(0)$ for all k. Hence, the function $V(k)$ measures the "disagreement" in agent values.

In the next lemma, we give a bound on the decrease of the agent disagreement $V(kB)$, which is linear in η and quadratic in n^{-1}. This bound is an immediate consequence of Lemma 5 in [13], stating that[3] under Assumptions 1 and 2, for all k with $V(kB) > 0$,

$$\frac{V(kB) - V((k+1)B)}{V(kB)} \geq \frac{\eta}{2n^2}.$$

This relation yields the following lemma.

Lemma 1: Let Assumptions 1 and 2 hold. Then, $V(k)$ is nonincreasing in k. Furthermore,

$$V((k+1)B) \leq \left(1 - \frac{\eta}{2n^2}\right) V(kB) \quad \text{for all } k \geq 0.$$

Using Lemma 1 we obtain the following result for the transition matrices $\Phi(k, s)$ of Eq. (3).

Corollary 1: Let Assumptions 1 and 2 hold. Then, for all i, j and all k, s with $k \geq s$, we have

$$\left| [\Phi(k,s)]_{ij} - \frac{1}{n} \right| \leq \left(1 - \frac{\eta}{4n^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 2}.$$

Proof: By Lemma 1, we have for all $k \geq s$,

$$V(kB) \leq \left(1 - \frac{\eta}{2n^2}\right)^{k-s} V(sB).$$

Let k and s be arbitrary with $k \geq s$, and let

$$\tau B \leq s < (\tau+1)B, \quad tB \leq k < (t+1)B,$$

with $\tau \leq t$. Hence, by the nonincreasing property of $V(k)$, we have

$$\begin{aligned} V(k) &\leq V(tB) \\ &\leq \left(1 - \frac{\eta}{2n^2}\right)^{t-\tau-1} V((\tau+1)B) \\ &\leq \left(1 - \frac{\eta}{2n^2}\right)^{t-\tau-1} V(s). \end{aligned}$$

Note that $k - s < (t-\tau)B + B$ implying that $\frac{k-s+1}{B} \leq t - \tau + 1$, where we used the fact that both sides of the inequality are integers. Therefore $\lceil \frac{k-s+1}{B} \rceil - 2 \leq t - \tau - 1$, and we have for all k and s with $k \geq s$,

$$V(k) \leq \left(1 - \frac{\eta}{2n^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 2} V(s). \tag{5}$$

By Eq. (4), we have $z(k+1) = A(k)z(k)$, and therefore $z(k+1) = \Phi(k,s)z(s)$ for all $k \geq s$. Letting $z(s) = e_i$ we obtain $z(k+1) = [\Phi(k,s)]^i$. Using the inequalities (5) and $V(e_i) \leq 1$, we obtain

$$V([\Phi(k,s)]^i) \leq \left(1 - \frac{\eta}{2n^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 2}.$$

[2]This update rule captures the averaging part of Eq. (2), as it operates on a particular component of the agent estimates, with the vector $z(k) \in \mathbb{R}^n$ representing the estimates of the different agents for that component.

[3]The assumptions in [13] are actually weaker.

The matrix $\Phi(k,s)$ is doubly stochastic, because it is the product of doubly stochastic matrices. Thus, the average entry of $[\Phi(k,s)]_i$ is $1/n$ implying that for all i and j,

$$\begin{aligned}\left([\Phi(k,s)]_{ij} - \frac{1}{n}\right)^2 &\leq V([\Phi(k,s)]^i) \\ &\leq \left(1 - \frac{\eta}{2n^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 2}.\end{aligned}$$

From the preceding relation and $\sqrt{1-\eta/(2n^2)} \leq 1 - \eta/(4n^2)$, we obtain

$$\left|[\Phi(k,s)]_{ij} - \frac{1}{n}\right| \leq \left(1 - \frac{\eta}{4n^2}\right)^{\lceil \frac{k-s+1}{B} \rceil - 2}.$$

■

B. Convergence time

We now study the convergence of the subgradient method (2) and obtain a convergence time bound. We assume the uniform boundedness of the set of subgradients of the cost functions f_i at all points[4]: for some scalar $L > 0$, we have for all $x \in \mathbb{R}^m$ and all i,

$$\|g\| \leq L \qquad \text{for all } g \in \partial f_i(x), \tag{6}$$

where $\partial f_i(x)$ is the set of all subgradients of f_i at x.

We define the time-averaged vectors $\hat{x}_i(k)$ of the iterates $x_i(k)$ generated by Eq. (2), i.e.,

$$\hat{x}_i(k) = \frac{1}{k}\sum_{h=0}^{k-1} x_i(h). \tag{7}$$

The use of these vectors allows us to bound the objective function improvement at every iteration; see [15], [14]. Under the subgradient boundedness assumption, we have the following result[5].

Theorem 2: Let Assumptions 1 and 2 hold, and assume that the set X^* of optimal solutions of problem (1) is nonempty. Let the sets of subgradients be bounded as in Eq. (6). Also, let the initial vectors $x_i(0)$ in Eq. (2) be such that $\max_{1\leq i\leq n}\|x_i(0)\| \leq \alpha L$. Then, the averages $\hat{x}_i(k)$ of the iterates obtained by the method (2) satisfy

$$\begin{aligned}f(\hat{x}_i(k)) &\leq f^* + \frac{n\,\text{dist}^2(y(0),X^*)}{2\alpha k} \\ &\quad + \frac{\alpha L^2 C}{2} + 2\alpha n L^2 C_1,\end{aligned}$$

where

$$C = 1 + 8nC_1,\ C_1 = 1 + \frac{nB}{\beta(1-\beta)},\ \beta = 1 - \frac{\eta}{4n^2}, \tag{8}$$

and $y(0) = (1/n)\sum_{i=1}^n x_i(0)$.

Proof: The proof follows from the forthcoming Theorem 3, as discussed at the end of Section III. ■

The convergence rate result in the preceding theorem improves that of Proposition 3 in [15], where an analogous estimate is shown with a worse value for the constant β. In particular, there the constant β in [15] is given by $\beta = 1 - \eta^{(n-1)B}$, and C_1 increases exponentially with n. As seen from Eq. (8), our new constants C and C_1 increases only polynomially with n indicating a much more favorable scaling as the network size increases. When α is fixed, the largest error is of the order of n^4, indicating that for high accuracy, the stepsize needs to be very small. However, our bound is for general convex functions and network topologies, and further improvements of the bound are possible for special classes of convex functions and special topologies.

III. Quantization effects

We next study the effects of quantization on the convergence properties of the subgradient method. In particular, we assume that each agent receives and sends only quantized estimates, i.e., vectors whose entries are integer multiples of $1/Q$. At time k, an agent receives quantized estimates $x_j^Q(k)$ from some of its neighbors and updates according to the following rule:

$$x_i^Q(k+1) = \left\lfloor \sum_{j=1}^n a_{ij}(k)x_j^Q(k) - \alpha\tilde{d}_i(k) \right\rfloor, \tag{9}$$

where $\tilde{d}_i(k)$ is a subgradient of f_i at $x_i^Q(k)$, and $\lfloor y \rfloor$ denotes the operation of (componentwise) rounding the entries of a vector y to the nearest multiple of $1/Q$. We also assume that the agents' initial estimates $x_j^Q(0)$ are quantized.

To analyze the proposed method, we find it useful to rewrite Eq. (9) as follows:

$$x_i^Q(k+1) = \sum_{j=1}^n a_{ij}(k)x_j^Q(k) - \alpha\tilde{d}_i(k) - e_i(k+1), \tag{10}$$

where the error vector $e_i(k+1)$ is given by

$$e_i(k+1) = \sum_{j=1}^n a_{ij}(k)x_j^Q(k) - \alpha\tilde{d}_i(k) - x_i^Q(k+1). \tag{11}$$

Thus, the method can be viewed as a subgradient method with external (possibly persistent) noise, represented by $e_i(k+1)$. Due to the rounding down to

[4]This assumption can be relaxed, see [15].

[5]The assumption $\max_{1\leq i\leq n}\|x_i(0)\| \leq \alpha L$ in this theorem is not essential. We use this assumption mainly to present a more compact expression for the bound on the convergence time. A bound that explicitly depends on $\|x_i(0)\|$ can be obtained by following a similar line of analysis

the nearest multiple of $1/Q$, the error vector $e_i(k+1)$ satisfies

$$0 \leq e_i(k+1) \leq \frac{1}{Q}\,\mathbf{1}, \qquad \text{for all } i \text{ and } k, \quad (12)$$

where the inequalities above hold componentwise.

Using the transition matrices $\Phi(k,s)$, we can rewrite the update equation (10) as

$$\begin{aligned} x_i^Q(k+1) &= \sum_{j=1}^{n}[\Phi(k,0)]_{ij}x_j^Q(0) \\ &\quad -\alpha\sum_{s=1}^{k}\sum_{j=1}^{n}[\Phi(k,s)]_{ij}\tilde{d}_j(s-1) \\ &\quad -\sum_{s=1}^{k}\sum_{j=1}^{n}[\Phi(k,s)]_{ij}e_j(s) \\ &\quad -\alpha\tilde{d}_i(k)-e_i(k+1). \qquad (13)\end{aligned}$$

In addition, we consider a related *stopped model*, where after some time $\tilde{k}$, *the agents cease computing subgradients* $\tilde{d}_j(k)$, *and also after time* $\tilde{k}+1$ stop quantizing (so that they can send and receive real numbers). Thus, in this stopped model, we have

$$\tilde{d}_i(k)=0 \text{ and } e_i(k+1)=0, \quad \text{for all } i \text{ and } k\geq\tilde{k}.$$

Let $\{\tilde{x}_i(k)\}$, $i=1,\ldots,n$ be the sequences generated by the stopped model, associated with a particular time $\tilde{k}$. In view of the preceding relation, we have for each i,

$$\tilde{x}_i(k)=x_i^Q(k) \qquad \text{for } k\leq\tilde{k},$$

and for $k\geq\tilde{k}+1$,

$$\begin{aligned} \tilde{x}_i(k) &= \sum_{j=1}^{n}[\Phi(k,0)]_{ij}x_j^Q(0) \\ &\quad -\alpha\sum_{s=1}^{\tilde{k}}\sum_{j=1}^{n}[\Phi(k,s)]_{ij}\tilde{d}_j(s-1) \\ &\quad -\sum_{s=1}^{\tilde{k}}\sum_{j=1}^{n}[\Phi(k,s)]_{ij}e_j(s). \qquad (14)\end{aligned}$$

Using the result of Corollary 1, we can show that the stopped process converges as $k\to\infty$. In particular, we have the following result.

Lemma 2: Let Assumptions 1 and 2 hold. Then, for any i and any $\tilde{k}\geq 0$, the sequence $\{\tilde{x}_i(k)\}$ generated by Eq. (14) converges and the limit vector does not depend on i, i.e.,

$$\lim_{k\to\infty}\tilde{x}_i(k)=y(\tilde{k}) \qquad \text{for all } i \text{ and } \tilde{k}.$$

Furthermore, for the limit sequence $y(k)$, we have:

(a) For all k,

$$y(k+1)=y(k)-\frac{\alpha}{n}\sum_{j=1}^{n}\tilde{d}_j(k)-\frac{1}{n}\sum_{j=1}^{n}e_j(k+1).$$

(b) When the subgradient norms $\|\tilde{d}_j(k)\|$ are uniformly bounded by some scalar L [cf. Eq. (6)] and the agents' initial values are such that $\max_j\|x_j^Q(0)\|\leq\alpha L$, then for all i and k, $\|x_i^Q(k)-y(k)\|\leq$

$$2\left(\alpha L+\frac{\sqrt{m}}{Q}\right)\left(1+\frac{nB}{\beta(1-\beta)}\right),$$

where $\beta=1-\frac{\eta}{4n^2}$ and m is the dimension of the vectors x_i^Q.

Proof: By Corollary 1, for any $s\geq 0$, the entries $[\Phi(k,s)]_{ij}$ converge to $1/n$, as $k\to\infty$. By letting $k\to\infty$ in Eq. (14), we see that the limit $\lim_{k\to\infty}\tilde{x}_i(k)$ exists and is independent of i. Denote this limit by $y(\tilde{k})$, and note that it is given by

$$\begin{aligned} y(\tilde{k}) &= \frac{1}{n}\sum_{j=1}^{n}x_j^Q(0)-\frac{\alpha}{n}\sum_{s=1}^{\tilde{k}}\sum_{j=1}^{n}\tilde{d}_j(s-1) \\ &\quad -\frac{1}{n}\sum_{s=1}^{\tilde{k}}\sum_{j=1}^{n}e_j(s). \qquad (15)\end{aligned}$$

From the preceding relation, applied to different values of $\tilde{k}$, we see that

$$y(k+1)=y(k)-\frac{\alpha}{n}\sum_{j=1}^{n}\tilde{d}_j(k)-\frac{1}{n}\sum_{j=1}^{n}e_j(k+1).$$

This establishes part (a).

Using the relations in Eqs. (13) and (15), and the subgradient boundedness, we obtain for all k,

$$\begin{aligned} \|x_i^Q(k)-y(k)\| &\leq \sum_{j=1}^{n}\left|[\Phi(k,0)]_{ij}-\frac{1}{n}\right|\|x_j^Q(0)\| \\ &\quad +\alpha L\sum_{s=1}^{k-1}\sum_{j=1}^{n}\left|[\Phi(k,s)]_{ij}-\frac{1}{n}\right| \\ &\quad +\sum_{s=1}^{k-1}\sum_{j=1}^{n}\left|[\Phi(k,s)]_{ij}-\frac{1}{n}\right|\|e_j(s)\| \\ &\quad +2\alpha L+\|e_i(k)\|+\frac{1}{n}\sum_{j=1}^{n}\|e_j(k)\|.\end{aligned}$$

By using Corollary 1, we have for all i and j, and any $k\geq s$,

$$\begin{aligned} \|x_i^Q(k)-y(k)\| &\leq \sum_{j=1}^{n}\beta^{\lceil\frac{k+1}{B}\rceil-2}\|x_j^Q(0)\| \\ &+ \alpha L\sum_{s=1}^{k-1}\sum_{j=1}^{n}\beta^{\lceil\frac{k-s+1}{B}\rceil-2} \\ &+ \sum_{s=1}^{k-1}\sum_{j=1}^{n}\beta^{\lceil\frac{k-s+1}{B}\rceil-2}\|e_j(s)\| \\ &+ 2\alpha L+\|e_i(k)\|+\frac{1}{n}\sum_{j=1}^{n}\|e_j(k)\|.\end{aligned}$$

Since $e_i(k) \le \mathbf{1}/Q$ [cf. Eq. (12)], we have

$$\|e_i(k)\| \le \frac{\sqrt{m}}{Q} \qquad \text{for all } i \text{ and } k.$$

From the preceding two relations, and the inequality $\max_j \|x_j^Q(0)\| \le \alpha L$, we obtain for all i and k,

$$\begin{aligned}\|x_i^Q(k) - y(k)\| &\le \alpha L n \beta^{\lceil \frac{k+1}{B} \rceil - 2} \\ &+ \alpha L n \sum_{s=1}^{k-1} \beta^{\lceil \frac{k-s+1}{B} \rceil - 2} + \frac{\sqrt{m}}{Q} n \sum_{s=1}^{k-1} \beta^{\lceil \frac{k-s+1}{B} \rceil - 2} \\ &+ 2\alpha L + 2\frac{\sqrt{m}}{Q}.\end{aligned}$$

By using $\sum_{s=0}^{k-1} \beta^{\lceil \frac{k-s+1}{B} \rceil - 2} \le \frac{1}{\beta} \sum_{r=0}^{\infty} \beta^{\lceil \frac{r+2}{B} \rceil - 1}$, and

$$\sum_{r=0}^{\infty} \beta^{\lceil \frac{r+2}{B} \rceil - 1} = \sum_{r=0}^{\infty} \beta^{\lceil \frac{r+2}{B} \rceil - 1} \le B \sum_{t=0}^{\infty} \beta^t = \frac{B}{1-\beta},$$

we finally obtain

$$\|x_i^Q(k) - y(k)\| \le 2\left(\alpha L + \frac{\sqrt{m}}{Q}\right)\left(1 + \frac{nB}{\beta(1-\beta)}\right).$$

■

According to part (a) of Lemma 2, the vectors $y(k)$ can be viewed as the iterates produced by the "fictitious" centralized algorithm:

$$y(k+1) = y(k) - \frac{\alpha}{n} \sum_{j=1}^{n} \tilde{d}_j(k) - \frac{1}{n} \sum_{j=1}^{n} e_j(k+1), \tag{16}$$

which is an approximate subgradient method with persistent noise: The direction $\sum_{j=1}^n \tilde{d}_j(k)$ is an approximate subgradient of the objective function f because each vector $\tilde{d}_j(k)$ is a subgradient of f_i at $x_i^Q(k)$ instead of at $y(k)$. The error term $(1/n)\sum_{j=1}^n e_j(k+1)$ can be viewed as the noise experienced by the whole system. The noise is persistent since the magnitudes of the errors $e_j(k)$ are non-diminishing.

We now focus on establishing an error bound for the function values at the points $y(k)$ of the stopped process of Eq. (16), starting with $y(0) = \frac{1}{n}\sum_{j=1}^n x_j^Q(0)$, and with the direction $\tilde{d}_j(k)$ being a subgradient of f_j at $x_j^Q(k)$ for all j and k. The process $y(k)$ is similar to the stopped process analyzed in [15], defined using $x_j(k)$ in place of $x_j^Q(k)$. Thus, using the same analysis as in [15] (see Lemma 5 therein), we can show the following basic result.

Lemma 3: Let Assumptions 1 and 2 hold, and assume that the set X^* of optimal solutions of problem (1) is nonempty. Let the sequence $\{y(k)\}$ be defined by Eq. (16), and the sequences $\{x_j^Q(k)\}$ for $j \in \{1, \ldots, n\}$ be generated by the quantized subgradient method (9). Also, assume that the subgradients are uniformly bounded as in Eq. (6), and that $\max_j \|x_j^Q(0)\| \le \alpha L$. Then, the average vectors $\hat{y}(k)$ defined as in Eq. (7), satisfy for all $k \ge 1$,

$$f(\hat{y}(k)) \le f^* + \frac{n\,\mathrm{dist}^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 \tilde{C}}{2},$$

where

$$\tilde{C} = 1 + \frac{8n\tilde{C}_1}{\alpha L},$$

$$\tilde{C}_1 = \left(\alpha L + \frac{\sqrt{m}}{Q}\right)\left(1 + \frac{nB}{\beta(1-\beta)}\right),$$

$\beta = 1 - \frac{\eta}{4n^2}$ and $y(0) = \frac{1}{n}\sum_{j=1}^n x_j^Q(0)$.

Proof: Using the same line of analysis as in the proof of Lemma 5 in [15], we can show that for all k,

$$\begin{aligned}\mathrm{dist}^2(y(k+1), X^*) &\le \mathrm{dist}^2(y(k), X^*) \\ &+ \frac{2\alpha}{n} \sum_{j=1}^{n} \left(\|\tilde{d}_j(k)\| + \|g_j(k)\|\right) \|y(k) - x_j^Q(k)\| \\ &- \frac{2\alpha}{n}[f(y(k)) - f^*] + \frac{\alpha^2}{n^2} \sum_{j=1}^{n} \|\tilde{d}_j(k)\|^2,\end{aligned}$$

where $g_j(k)$ is a subgradient of f_j at $y(k)$. By using the subgradient boundedness, we further obtain

$$\begin{aligned}\mathrm{dist}^2(y(k+1), X^*) &\le \mathrm{dist}^2(y(k), X^*) \\ &+ \frac{4\alpha L}{n} \sum_{j=1}^{n} \|y(k) - x_j^Q(k)\| \\ &- \frac{2\alpha}{n}[f(y(k)) - f^*] + \frac{\alpha^2 L^2}{n}.\end{aligned}$$

By using Lemma 2(b), we have

$$\begin{aligned}\mathrm{dist}^2(y(k+1), X^*) &\le \mathrm{dist}^2(y(k), X^*) \\ &+ 8\alpha L \tilde{C}_1 - \frac{2\alpha}{n}[f(y(k)) - f^*] + \frac{\alpha^2 L^2}{n},\end{aligned}$$

where $\tilde{C}_1 = \left(\alpha L + \frac{\sqrt{m}}{Q}\right)\left(1 + \frac{nB}{\beta(1-\beta)}\right)$. Therefore,

$$\begin{aligned}f(y(k)) &\le f^* + \frac{\alpha L^2}{2} + 4nL\tilde{C}_1 \\ &+ \frac{n}{2\alpha}\left(\mathrm{dist}^2(y(k), X^*) - \mathrm{dist}^2(y(k+1), X^*)\right),\end{aligned}$$

and by regrouping the terms and introducing $\tilde{C} = 1 + \frac{8n\tilde{C}_1}{\alpha L}$, we have for all k,

$$\begin{aligned}f(y(k)) &\le f^* + \frac{\alpha L^2 \tilde{C}}{2} \\ &+ \frac{n}{2\alpha}\left(\mathrm{dist}^2(y(k), X^*) - \mathrm{dist}^2(y(k+1), X^*)\right).\end{aligned}$$

By adding these inequalities for different values of k, and by using the convexity of f, we obtain the desired inequality. ■

Assuming that the agents can store real values (infinitely many bits), we consider the time-average of the iterates $x_i^Q(k)$, defined by

$$\hat{x}_i^Q(k) = \frac{1}{k}\sum_{h=0}^{k-1} x_i^Q(h) \qquad \text{for } k \geq 1.$$

Using Lemma 3, we have the following result.

Theorem 3: Under the same assumptions as in Lemma 3, the averages $\hat{x}_i^Q(k)$ of the iterates obtained by the method (9) satisfy, for all i,

$$f(\hat{x}_i^Q(k)) \leq f^* + \frac{n\,\mathrm{dist}^2(y(0), X^*)}{2\alpha k} + \frac{\alpha L^2 \tilde{C}}{2} + 2nL\tilde{C}_1,$$

where $\tilde{C}$, $\tilde{C}_1$, and $y(0)$ are as in Lemma 3.

Proof: By the convexity of the functions f_j, we have, for any i and k,

$$f(\hat{x}_i^Q(k)) \leq f(\hat{y}(k)) + \sum_{j=1}^{n} g_{ij}(k)'(\hat{x}_i^Q(k) - \hat{y}(k)),$$

where $g_{ij}(k)$ is a subgradient of f_j at $\hat{x}_i^Q(k)$. Then, by using the boundedness of the subgradients and Lemma 2(b), we obtain for all i and k,

$$f(\hat{x}_i^Q(k)) \leq f(\hat{y}(k)) + 2nL\tilde{C}_1,$$

with $\tilde{C}_1 = \left(\alpha L + \frac{\sqrt{m}}{Q}\right)\left(1 + \frac{nB}{\beta(1-\beta)}\right)$. The result follows by using Lemma 3. ■

The result of Theorem 2 follows from Theorem 3 by letting the quantization level Q be increasingly finer (i.e., $Q \to \infty$). Specifically, when $Q \to \infty$, the constants $\tilde{C}_1$ and $\tilde{C}$ of Theorem 3 satisfy

$$\lim_{Q\to\infty} \tilde{C}_1 = \alpha L \left(1 + \frac{nB}{\beta(1-\beta)}\right) = \alpha L C_1,$$

$$\lim_{Q\to\infty} \tilde{C} = 1 + \frac{8n}{\alpha L} \lim_{Q\to\infty} \tilde{C}_1 = 1 + 8nC_1,$$

with $C_1 = 1 + \frac{nB}{\beta(1-\beta)}$. Thus, for the error term of Theorem 3, we have

$$\lim_{Q\to\infty}\left(\frac{\alpha L^2 \tilde{C}}{2} + 2nL\tilde{C}_1\right) = \frac{\alpha L^2}{2} C + 2n\alpha L^2 C_1$$

where $C = 1 + 8nC_1$ and $C_1 = 1 + \frac{nB}{\beta(1-\beta)}$. Hence, in the limit as $Q \to \infty$, the estimate in Theorem 3 yields the estimate in Theorem 2.

IV. Conclusions

We studied distributed subgradient methods for convex optimization problems that arise in networks of agents connected through a time-varying topology. We first considered an algorithm for the case where agents can exchange and store continuous values, and proved a bound on the convergence rate. We next studied the algorithm under the additional constraint that agents can only send and receive quantized values. We showed that our algorithm guarantees convergence of the agent values to the optimal objective value within some error. Our bound on the error highlights the dependence on the number of quantization levels, and the polynomial dependence on the number n of agents. Future work includes studying the effects of other quantization schemes and of noise in the agents' estimates.

References

[1] D.P. Bertsekas, *Nonlinear programming*, Athena Scientific, Cambridge, Massachusetts, 1999.

[2] D.P. Bertsekas, A. Nedić, and A.E. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Cambridge, Massachusetts, 2003.

[3] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont, MA, 1997.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Gossip algorithms: Design, analysis, and applications*, Proceedings of IEEE INFOCOM, 2005.

[5] M. Cao, D. A. Spielman, and A. S. Morse, *A lower bound on convergence of a distributed network consensus algorithm*, Proceedings of IEEE CDC, 2005.

[6] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and S. Zampieri, *Average consensus on networks with transmission noise or quantization*, Proc. of European Control Conference, 2007.

[7] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in coordinated consensus problems*, Proceedings of IEEE ACC, 2006.

[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Springer-Verlag, Berlin, 1996.

[9] A. Jadbabaie, J. Lin, and S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Transactions on Automatic Control **48** (2003), no. 6, 988–1001.

[10] A. Kashyap, T. Basar, and R. Srikant, *Quantized consensus*, Automatica **43** (2007), no. 7, 1192–1203.

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

[12] S. Low and D.E. Lapsley, *Optimization flow control, I: Basic algorithm and convergence*, IEEE/ACM Transactions on Networking **7** (1999), no. 6, 861–874.

[13] A. Nedić, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis, *On distributed averaging algorithms and quantization effects*, Tech. Report LIDS 2778, MIT, 2007.

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

[15] ———, *Distributed subgradient methods for multi-agent optimization*, IEEE Transactions on Automatic Control, forthcoming.

[16] R. Olfati-Saber and R.M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control **49** (2004), no. 9, 1520–1533.

[17] A. Olshevsky and J.N. Tsitsiklis, *Convergence rates in distributed consensus and averaging*, Proceedings of IEEE CDC, 2006.

[18] ———, *Convergence speed in distributed consensus and averaging*, forthcoming in SIAM Journal on Control and Optimization, 2008.

[19] B.T. Polyak, *A general method for solving extremum problems*, Soviet Math. Doklady **8** (1967), no. 3, 593–597.

[20] ———, *Introduction to optimisation*, Optimization Software, Inc., New York, 1987.
[21] S. S. Ram, A. Nedić, and V. V. Veeravalli, *Stochastic incremental gradient descent for estimation in sensor networks*, Proc. of Asilomar, 2007.
[22] N.Z. Shor, *Minimization methods for nondifferentiable functions*, Translated from Russian by K.C. Kiwiel and A. Ruszczynski, Springer, Berlin, 1985.
[23] R. Srikant, *Mathematics of Internet congestion control*, Birkhauser, 2004.
[24] J.N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, MIT, 1984.
[25] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.
[26] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Schochet, *Novel type of phase transitions in a system of self-driven particles*, Physical Review Letters **75** (1995), no. 6, 1226–1229.
[27] L. Xiao, S. Boyd, and S. Lall, *A space-time diffusion scheme for peer-to-peer least-squares estimation*, Proc. of IPSN, 2006.