DISTRIBUTED OPTIMIZATION IN AN ENERGY-CONSTRAINED NETWORK USING A DIGITAL COMMUNICATION SCHEME

Alireza Razavi and Zhi-Quan Luo

Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455
Email: {raza0007, luozq}@umn.edu

ABSTRACT

We consider a distributed optimization problem where n nodes, S_l, $l \in \{1, \ldots, n\}$, wish to minimize a common strongly convex function $f(\mathbf{x})$, $\mathbf{x} = [x_1, \ldots, x_n]^T$, and suppose that node S_l only has control of variable x_l. The nodes locally update their respective variables and periodically exchange their values over noisy channels. Previous studies of this problem have mainly focused on the convergence issue and the analysis of convergence rate. In this work, we focus on the communication energy and study its impact on convergence. In particular, we study the minimum amount of communication energy required for nodes to obtain an ϵ-minimizer of $f(\mathbf{x})$ in the mean square sense. In an earlier work, we considered analog communication schemes and proved that the communication energy must grow at the rate of $\Omega(\epsilon^{-1})$ to obtain an ϵ-minimizer of a convex quadratic function. In this paper, we consider digital communication schemes and propose a distributed algorithm which only requires communication energy of $\mathcal{O}\left(\left(\log \epsilon^{-1}\right)^3\right)$ to obtain an ϵ-minimizer of $f(\mathbf{x})$. Furthermore, the algorithm provided herein converges linearly. Thus, distributed optimization with digital communication schemes is significantly more energy efficient than with analog communication schemes.

Index Terms— Distributed optimization, Sensor networks, Energy constraint, Convergence

1. INTRODUCTION

Consider a network of n nodes which collaborate to minimize a cost $f(\mathbf{x})$, $\mathbf{x} = [x_1, \ldots, x_n]^T$, where x_l is a local (vector) variable controlled by node S_l. Each node can perform local computation and exchange messages with a set of predefined neighbors through orthogonal noisy channels. Moreover, we assume $f(\mathbf{x})$ has a certain *"local structure"* in the sense that its partial derivative with respect to x_l only depends on the local variables at node S_l and its neighbors. A distributed optimization problem of this kind arises naturally in sensor network applications. For example, in the sensor localization problem, we are given the locations of anchor nodes and distance measurements between certain neighbor nodes in the network. The goal is to estimate the locations of all sensors in the network by distributed minimization of a cost function $f(\mathbf{x})$ defined by the L_p norm of distance errors [2]. In this context, x_l is the location of sensor S_l and is to be estimated by S_l. To minimize $f(\mathbf{x})$, sensor S_l periodically updates its local variable, x_l, and exchanges information with neighbor nodes through orthogonal noisy channels. A special feature of this problem is the fact that nodes are usually battery operated and hence energy-constrained. Note that energy of each node is consumed for various operations including local computation and inter-sensor communication, with the latter being the dominant part. This motivates us to study the minimum amount of communication energy required for distributed optimization.

Energy consumption has not been a consideration of algorithm design in classical distributed optimization [1]. Even recent studies of distributed optimization in the context of sensor networks [7, 4] have mainly focused on convergence issues such as convergence criteria and convergence rate. To the best of our knowledge, the most relevant work to this paper is [6] which studied the minimum number of bits that must be exchanged between two nodes in order to find an ϵ-minimizer of f. However, unlike our current work, the communication channel is assumed distortion-less in [6], and there was no effort to characterize minimum energy consumption.

Recently, we considered an analog communication scheme for this distributed optimization problem and proved that the communication energy must grow at the rate of $\Omega(\epsilon^{-1})$ in order to obtain an ϵ-minimizer of convex quadratic $f(\mathbf{x})$ [5]. In this paper, instead, we consider digital communication schemes for a wider class of cost functions, strongly convex function, and propose a distributed algorithm which requires $\mathcal{O}\left(\left(\log \epsilon^{-1}\right)^3\right)$ communication energy to obtain an ϵ-minimizer of $f(\mathbf{x})$. Furthermore, the algorithm provided herein converges linearly to the optimal solution as compared to our previous algorithm which has a sub-linear convergence rate. Thus, digital communication schemes are far more energy efficient than analog communication schemes for distributed optimization.

2. ALGORITHM FRAMEWORK

Let $\mathcal{F}_{SC,M,L}$ ('strongly convex functions') be a set of continuously differentiable function $f(\mathbf{x})$ with the properties

$$L\,\|\mathbf{x}-\mathbf{y}\|^2 \leq \langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x}-\mathbf{y}\rangle \leq M\,L\,\|\mathbf{x}-\mathbf{y}\|^2\,, \quad (1)$$

where $\nabla f(\mathbf{x})$ is gradient vector of function $f(\mathbf{x})$ at point $\mathbf{x}$, and M and L are positive numbers.

Consider a distributed optimization problem where n nodes, S_l, $l \in \{1, \ldots, n\}$, jointly minimize a common cost function $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}, \mathbf{x} = [x_1, \ldots, x_n]^T$. Node S_l only has control of variable x_l and has ability to compute the partial derivative of the cost function with respect to its local variable. Furthermore, we assume that the local variable $x_l, l \in \{1, \ldots, n\}$ has a finite range and is bounded to $[0, 1]$.

We assume that nodes communicate through orthogonal time-invariant noisy channels. The communication channel between nodes

This work is supported by the USDOD Army, grant number W911NF-05-1-0567 and by Digital Technology Center, University of Minnesota.

S_l and S_j is corrupted by Additive White Gaussian Noise (AWGN) with power spectral density $N_0/2$:

$$\widehat{m}_{l,j} = d_{l,j}^{\kappa/2} m_l + v_{l,j},$$

where $\hat{m}_{l,j}$ is the received message at node S_j from node S_l, and $v_{l,j}$ is the AWGN. The signal power received at node S_j is assumed to be inversely proportional to $d_{l,j}^{\kappa}$, where $d_{l,j}$ is the distance between nodes S_j and S_l, and κ is the path loss exponent. We assume that energy required for transmission of m_l is proportional to the number of bits in the message (b_l). This is the case e.g., if nodes use M-QAM or M-PSK modulation to transmit messages. For example, if M-QAM is used, the energy per bit $W_l(P_b)$ is [3]:

$$\begin{aligned} W_l(P_b) &= \frac{4N_f \max_j N_{l,j} d_{l,j}^{\kappa} G_0(2^s-1)}{3s} \log\left(\frac{4(1-2^{-\frac{s}{2}})}{sP_b}\right) \\ &= w_l \log\left(\frac{4(1-2^{-\frac{s}{2}})}{sP_b}\right), \end{aligned}$$

where s is the number of bits per M-QAM symbol, N_f is the receiver noise figure, $N_{l,j}$ is the power spectrum density of channel noise between nodes S_l and S_j, G_0 is the system constant defined as in [3], and P_b is the required bit error probability. Therefore, the total communication energy is

$$E_{com}(t) = \sum_{i=1}^{t} \sum_{l=1}^{n} W_l(P_b(i)) b_l(i).$$

This paper aims to study the minimum communication energy required to obtain an ϵ-minimizer of $f(\mathbf{x})$ in the mean square sense. A point $\mathbf{x}$ is an ϵ-minimizer of $f(\mathbf{x})$ in the mean square sense if

$$\mathbb{E}[\|\mathbf{x}-\mathbf{x}^*\|^2] \le \epsilon, \ \mathbf{x}^* = \arg\min_x f(\mathbf{x}).$$

Here, we introduce a distributed algorithm in which node S_l iteratively updates its local variable x_l and tracks the variables of its neighbors. The algorithm consists of two parts: A digital communication scheme and a local computation scheme at each node.

A. Communication scheme: After each local update, node S_l should relay its information to the other nodes. One way is to directly send the quantized version of the updated value of its local variable, $x_i^i(t+1)$, which requires transmitting more bits, and thus consuming more energy, as algorithm proceeds. An alternative way is to send the quantized version the incremental value, $x_i^i(t+1) - x_i^i(t)$, in which case requires transmitting a constant number of bits at each iteration but may not guarantee the convergence of the algorithm due to communication error. In general, we can consider a linear messaging scheme where the transmitted signal, $m_l(t)$, is given as

$$m_l(t) = \mathcal{Q}(x_l^l(t) - \gamma(t)x_l^l(t-1)) = [\mathrm{sgn}_l(t), \mathrm{amp}_l(t)] \quad (2)$$

where $\mathrm{sgn}_l(t)$ represents the sign of $x_l^l(t) - \gamma(t)x_l^l(t-1)$, and $\mathrm{amp}_l(t)$ is the binary representation of integer part of $2^{R(t)}|x_l^l(t) - \gamma(t)x_l^l(t-1)|$. The resolution of quantizer, $R(t)$, is a design parameter and will be determined later. Receiver node S_j first detects $\widehat{m}_l(t) = [\widehat{\mathrm{sgn}}_l(t), \widehat{\mathrm{amp}}_l(t)]$ and then reconstructs x_l^j as following ($x_l^j(0) = 0$)

$$x_l^j(t) = \gamma(t)x_l^j(t-1) + \widehat{\mathrm{sgn}}_l(t)2^{-R(t)}\widehat{\mathrm{amp}}_l(t). \quad (3)$$

Notice that variable $x_l^j(t)$ is a noisy copy of node S_l's variable $x_l^l(t)$ at node S_j. Define $n_{Q,l}(t)$, and $e_{l,j}(t)$ as quantization noise at node S_l at iteration t, and communication error due to channel noise between node S_l and S_j at iteration t, respectively.

$$\begin{aligned} n_{Q,l}(t) &= \mathrm{sgn}_l(t)2^{-R(t)}\mathrm{amp}_l(t) - \Big(x_l^l(t) - \gamma(t)x_l^l(t-)\Big) \\ e_{l,j}(t) &= \widehat{\mathrm{sgn}}_l(t)2^{-R(t)}\widehat{\mathrm{amp}}_l(t) - \mathrm{sgn}_l(t)2^{-R(t)}\mathrm{amp}_l(t) \end{aligned}$$

Then the variable $x_l^j(t)$ is given as

$$\begin{aligned} x_l^j(t) &= \gamma(t)x_l^j(t-1) + x_l^l(t) - \gamma(t)x_l^l(t-1) + n_{Q,l}(t) + e_{l,j}(t) \\ &= x_l^l(t) + \gamma(t)\Big(x_l^j(t-1) - x_l^l(t-1)\Big) + n_{Q,l}(t) + e_{l,j}(t) \\ &= x_l^l(t) + \sum_{i=1}^{t}\prod_{k=i+1}^{t}\gamma(k)\left(n_{Q,l}(i) + e_{l,j}(i)\right). \end{aligned} \quad (4)$$

B. Local computation scheme: Optimization algorithms in the presence of noise can be performed based on the gradient projection algorithm proposed in [6]. One iteration of this algorithm is given as

$$\mathbf{x}(t) = [\mathbf{x}(t-1) - a\,\mathbf{g}(\mathbf{x}(t-1))]_+, \mathbf{x}(0) = 0. \quad (5)$$

where $[x]_+$ is the projection of x on $[0,1]$, a is a constant positive step size, and $\mathbf{g}(\mathbf{x}(t))$ is a noisy version of the gradient vector $\nabla f(\mathbf{x}(t))$. For $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}$, it has been proven in [6] that if $\mathbf{g}(\mathbf{x}(t))$ satisfies

$$\|\mathbf{g}(\mathbf{x}(t)) - \nabla f(\mathbf{x}(t))\| \le n^{1/2}\alpha^t, \quad t = 1, \ldots, \quad (6)$$

for $\alpha < 1$ sufficiently close to 1, then the sequence $\{\mathbf{x}(t)\}$ generated by the gradient projection algorithm (5) converges linearly to the optimal point.

We consider a distributed implementation of the gradient projection algorithm whereby $S_j, j \in \{1, \ldots, n\}$ uses the noisy copy of its neighbors variables to estimate $g_j(\mathbf{x})$, the partial derivative of $f(\mathbf{x})$ with respect to its local variable x_j, and to update x_j^j as ($x_j^j(0) = 0$)

$$x_j^j(t) = \left[x_j^j(t-1) - \frac{1}{LM} g_j\left(x_1^j(t), \ldots, x_n^j(t)\right)\right]_+. \quad (7)$$

Here, we have chosen $a = \frac{1}{ML}$. In the next section, we first derive a sufficient condition on the coefficient $\gamma(t)$, the resolution $R(t)$, and probability of bit error $P_b(t)$ to obtain an ϵ-minimizer of $f(\mathbf{x})$. We then bound the total communication energy.

3. CONVERGENCE CONDITION AND COMMUNICATION ENERGY

Let $\mathbf{x}(t) = [x_1^1(t), \ldots, x_n^n(t)]^T$ and define $\mathbf{g}(\mathbf{x}(t))$ as a vector of the partial derivatives of $f(\mathbf{x})$ with respect to the local variables which are computed locally at each node using the noisy replicas of neighbors variables:

$$\mathbf{g}(\mathbf{x}(t)) = [g_1\left(x_1^1(t), \ldots, x_n^1(t)\right), \ldots, g_n\left(x_1^n(t), \ldots, x_n^n(t)\right)]^T.$$

Then, the distributed computation (7) can be expressed as the gradient projection algorithm

$$\mathbf{x}(t) = \left[\mathbf{x}(t-1) - \frac{1}{LM}\mathbf{g}(\mathbf{x}(t-1))\right]_+. \quad (8)$$

Notice that the convergence condition (6) does not satisfy for every realization of channel noise. Therefore, the sequence $\{\mathbf{x}(t)\}$ generated by (8) might not converge to the optimal point. However, under a modified assumption, the sequence $\{\mathbf{x}(t)\}$ converges to the optimal point $\mathbf{x}^*$ in the mean squared sense using the following lemma:

Lemma 1 *For $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}$, if $\mathbf{g}(\mathbf{x}(t))$ satisfies*

$$\mathbb{E}\left[\|\mathbf{g}(\mathbf{x}(t)) - \nabla f(\mathbf{x}(t))\|^2\right] \le n\alpha^{2t}, \quad t = 1, \ldots, \tag{9}$$

where

$$(1 - \frac{1}{M}) + 3\sqrt{(\frac{2}{LA})} \le \alpha^2 < 1, \tag{10}$$

then the sequence $\{\mathbf{x}(t)\}$ generated by the gradient projection algorithm (8) converges linearly to the optimal point $\mathbf{x}^*$ in the mean squared sense such that

$$\mathbb{E}\left[\|\mathbf{x}(t) - \mathbf{x}^*\|^2\right] \le \frac{2A}{L} n\alpha^{2t}.$$

where

The proof of Lemma **1** is similar to the proof of proposition 5.1 in [6] and omitted here for lack of space. We use Lemma **1** to show that the proposed algorithm generates an ϵ-minimizer of $f(\mathbf{x})$ if the design parameters are chosen as

$$\begin{aligned} \gamma(t) &= \alpha, & (11)\\ R(t) &= \left\lceil \log_2\left(\sqrt{\frac{2ML}{1-\alpha^2}}\alpha^{-2t}\right)\right\rceil, & (12)\\ P_b(t) &= \frac{3\left(1-\alpha^2\right)}{8LM}\alpha^{4t}, & (13) \end{aligned}$$

where $\lceil x \rceil$ is the smallest integer number greater than x, α is a positive constant less than 1, and $\log_y(x)$ is the logarithm base y of x.

Theorem 1 *If the design parameters, $\gamma(t)$, $R(t)$, and $P_b(t)$ satisfy (11)-(13), then the distributed algorithm described by* (2), (3), *and* (7) *obtains an ϵ-minimizer of $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}$ in the mean square sense. Moreover, the communication energy to obtain an ϵ-minimizer of $f(\mathbf{x})$, $\mathbf{x}(t_\epsilon)$, is*

$$E_{com}(t_\epsilon) = \mathcal{O}\left(\left(\log \epsilon^{-1}\right)^3\right).$$

Proof: We first show that the convergence condition (9) of Lemma **1** holds. Therefore, the proposed algorithm obtains an ϵ-minimizer of $f(\mathbf{x})$.

For $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}$, It follows form (1) that

$$\|\nabla f(\mathbf{x}) - \nabla f(\hat{\mathbf{x}})\| \le M\,L\,\|\mathbf{x} - \hat{\mathbf{x}}\|.$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{g}(\mathbf{x}(t)) - \nabla f(\mathbf{x}(t))\|^2\right] &\le ML\mathbb{E}\left[\sum_{l=1}^{n}\left(x_l^J(t) - x_l^l(t)\right)^2\right] \\ &= ML\sum_{l=1}^{n}\mathbb{E}\left[\left(x_l^J(t) - x_l^l(t)\right)^2\right], \quad (14) \end{aligned}$$

where J is given as

$$J = \arg\max_j \left\|\left[x_1^j(t) - x_1^1(t), \ldots, x_n^j(t) - x_n^n(t)\right]^T\right\|^2.$$

Substitute the difference between variable x_l^l and its noisy copy at node S_J, x_l^J, from equation (4), and use the fact that the quantization noise and communication error are uncorrelated to obtain

$$\begin{aligned} \mathbb{E}\left[\left(x_l^J(t) - x_l^l(t)\right)^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{t}\prod_{k=i+1}^{t}\gamma(k)(n_{Q,l}(i) + e_{l,J}(i))\right)^2\right] \\ &= \sum_{i=1}^{t}\prod_{k=i+1}^{t}\gamma^2(k)\left(\mathbb{E}\left[n_{Q,l}^2(i)\right] + \mathbb{E}\left[e_{l,J}^2(i)\right]\right). \quad (15) \end{aligned}$$

The power of quantization noise $\mathbb{E}\left[n_{Q,l}^2(i)\right]$ is upper-bounded by

$$\mathbb{E}\left[n_{Q,l}^2(i)\right] \le 2^{-2R(i)}, \tag{16}$$

where $R(i)$ is resolution of quantizer. Notice that the number of transmitted bits at iteration i is bounded by $R(i)$. Therefore, the power of communication error $\mathbb{E}\left[e_{l,J}^2(i)\right]$ is bounded by

$$\mathbb{E}\left[e_{l,J}^2(i)\right] \le \sum_{q=0}^{R(i)} P_b(i)2^{-2q} \le \frac{4P_b(i)}{3}. \tag{17}$$

Substitute (15)-(17) into (14) to derive

$$\begin{aligned} &\mathbb{E}\|\mathbf{g}(\mathbf{x}(t)) - \nabla f(\mathbf{x}(t))\|^2 \\ \le\;& nML\sum_{i=1}^{t}\prod_{k=i+1}^{t}\gamma^2(k)\left(2^{-2R(i)} + \frac{4P_b(i)}{3}\right). \end{aligned}$$

It follows from Lemma **1** that the proposed algorithm obtains an ϵ-minimizer of $f(\mathbf{x})$ if

$$\sum_{i=1}^{t}\frac{\left(2^{-2R(i)} + \frac{4P_b(i)}{3}\right)}{\prod_{k=1}^{i}\gamma^2(k)} \le \frac{\alpha^{2t}}{ML\prod_{k=1}^{t}\gamma^2(k)}, t = 1, \ldots. \tag{18}$$

Notice the left-hand side of the inequality (18) is a non-decreasing function of iteration number t. Hence, the right-hand of the inequality (18) should also be a non-decreasing function of iteration number. Therefore, we have $\alpha^2/\gamma^2(t) \ge 1$, and thus $\gamma(t) \le \alpha$. Let $\gamma(t)\alpha$. Then, to prove the convergence of the proposed algorithm, it is enough to show that

$$\begin{aligned} \sum_{i=1}^{t}\alpha^{-2i}2^{-2R(i)} &\le \frac{1}{2ML}, & (19)\\ \sum_{i=1}^{t}\alpha^{-2i}P_b(i) &\le \frac{3}{8ML}. & (20) \end{aligned}$$

Recall the resolution $R(i)$ from (12) and notice that $\alpha < 1$. Therefore, the inequality (19) holds as

$$\sum_{i=1}^{t}\alpha^{-2i}2^{-2R(i)} \le \frac{1-\alpha^2}{2LM}\sum_{i=1}^{t}\alpha^{2i} \le \frac{1}{2LM}.$$

Similarly, the inequality (20) holds for the choice of the bit error probability $P_b(i)$ at (13). Therefore, the convergence condition (9) is satisfied and the algorithm converges linearly to the optimal point in the mean squared sense (Lemma **1**). In particular, the algorithm obtains an ϵ-minimizer of $f(\mathbf{x})$ at iteration t_ϵ, if

$$t_\epsilon \ge \left\lceil \frac{1}{2}\log_\alpha\left(\frac{L\epsilon}{2An}\right)\right\rceil. \tag{21}$$

So far, we have shown that the proposed algorithm obtains an ϵ-minimizer of $f(\mathbf{x})$. Next, we derive the communication energy in terms of ϵ. Recall that the total communication energy is

$$E_{com}(t_\epsilon) = \sum_{i=1}^{t_\epsilon}\sum_{l=1}^{n} W_l(P_b(i))b_l(i).$$

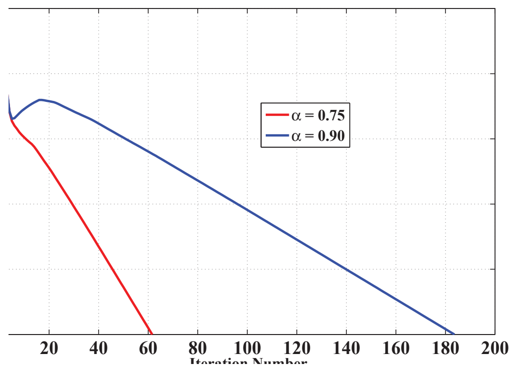

Fig. 1. Mean squared error

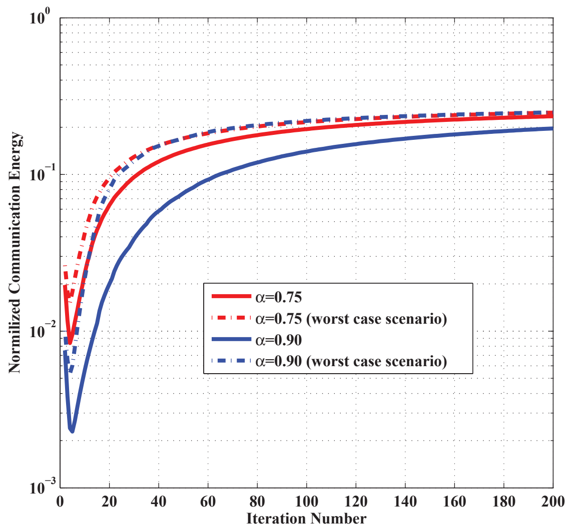

Fig. 2. Normalized communication energy

Notice that the communication energy per bit is

$$W_l(P_b(i)) = w_l \log\left(\frac{4(1-2^{-\frac{s}{2}})}{sP_b(i)}\right) = 4w_l \log\alpha^{-1} i + \mathcal{O}(1), \quad (22)$$

and the number of transmitted bits at iteration i $b_l(i)$ is bounded by $R(i)+1$. Substitute the resolution $R(i)$ from (12), the communication energy per bit $W_l(P_b(i))$ from (22) to obtain

$$E_{com}(t_\epsilon) \leq \sum_{i=1}^{t_\epsilon}\left(c_1 i^2 + \mathcal{O}(i)\right) = \frac{c_1}{3}t_\epsilon^3 + \mathcal{O}(t_\epsilon^2), \quad (23)$$

where constant c_1 is equal to $8n \max_l w_l \log\alpha^{-1}\log_2\alpha^{-1}$.

Replace the iteration number from (21) into (23) to derive

$$E_{com}(t_\epsilon) \quad = \quad c_2\left(\log\epsilon^{-1}\right)^3 + \mathcal{O}\left(\log\epsilon^{-1}\right)^2,$$

where constant constant c_2 is equal to $\frac{c_1}{24\left(\log\alpha^{-1}\right)^3}$. The theorem has been proven. ■

4. SIMULATION RESULTS

To illustrate the concept, we consider a simple example where 10 nodes minimize a quadratic convex function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x}+\mathbf{c}$, where $\mathbf{Q} \in R^{10\times10}$, $\mathbf{Q} \succ 0$. Notice that $f(\mathbf{x}) \in \mathcal{F}_{SC,M,L}$ where M and L are the condition number and minimum eigenvalue of matrix $\mathbf{Q}$, respectively. We generate a random matrix $\mathbf{A}$ such that $f(\mathbf{x}) \in \mathcal{F}_{SC,2,1}$, and select vector $\mathbf{b}$ such that $f(\mathbf{x})$ has a minimum point at $\mathbf{x} = [0.5,\ldots,0.5]^T$. Since all coefficients w_l are scaled by a common factor. In simulation, w_l are taken to be maximum of channel path loss $w_l = \max_j d_{l,j}^\kappa$, $\kappa = 2$. The mean square error for different value of α averaged over 10 runs is shown in Figure 1. This figure confirms that the algorithm converges linearly to the optimal point. Notice that the choice of α should satisfy (10). In particular, $\alpha \geq \sqrt{1-1/M} = \sqrt{2}/2$. Figure 2 shows that the communication energy required to obtain ϵ-minimizer, normalized by $c_2\left(\log\epsilon^{-1}\right)^3$, is bounded. This figure also shows the normalized communication energy for the worst case scenario where $R(i)+1$ number of bits is transmitted by each node at i^{th} iteration. These results agree with our theoretical analysis.

5. CONCLUSIONS AND FUTURE WORK

We studied the problem of distributed optimization of a strongly convex function in an energy-constrained network. We considered a class of distributed gradient projection algorithm implemented using certain digital messaging schemes. We proposed an algorithm which requires communication energy of order $\mathcal{O}\left(\left(\log\epsilon^{-1}\right)^3\right)$ to obtain an ϵ-minimizer. Furthermore, this algorithm converges linearly to the optimal solution. Our numerical simulations confirmed our theoretical analysis.

6. REFERENCES

[1] D. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[2] M. W. Carter, H. H. Jin, M. A. Saunders, and Y. Ye, "Spaseloc: An adaptive subproblem algorithm for scalable wireless sensor network localization," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1102–1128, 2006.

[3] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[4] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," in *Proceedings of 45th IEEE Conference on Decision and Control*, 2006, pp. 3387–3392.

[5] A. Razavi and Z.-Q. Luo, "Distributed optimization in an energy-constrained network," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 2007, pp. III–189–III–192.

[6] J. N. Tsitsiklis and Z. Luo, "Communication complexity of convex optimization," *Journal of Complexity*, vol. 3, no. 3, pp. 231–243, 1987.

[7] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of Optimization Theory and Applications*, vol. 129, no. 3, pp. 469–488, 2006.