

QUANTIZED CONSENSUS ADMM FOR MULTI-AGENT DISTRIBUTED OPTIMIZATION

Shengyu Zhu*, Mingyi Hong†, and Biao Chen*

* Department of EECS, Syracuse University, Syracuse, NY 13244, USA

† Department of IMSE, Iowa State University, Ames, IA 50011, USA

ABSTRACT

This paper considers multi-agent distributed optimization with quantized communication which is needed when inter-agent communications are subject to finite capacity and other practical constraints. To minimize the global objective formed by a sum of local convex functions, we develop a quantized distributed algorithm based on the alternating direction method of multipliers (ADMM). Under certain convexity assumptions, it is shown that the proposed algorithm converges to a consensus within $\log_{1+\eta} \Omega$ iterations, where $\eta > 0$ depends on the network topology and the local objectives, and Ω is a polynomial fraction depending on the quantization resolution, the distance between initial and optimal variable values, the local objectives, and the network topology. We also obtain a tight upper bound on the consensus error which does not depend on the size of the network.

Index Terms— Multi-agent distributed optimization, quantization, alternating direction method of multipliers (ADMM), linear convergence.

1. INTRODUCTION

There has been much research interest in distributed optimization due to recent advances in networked multi-agent systems [1, 2]. For example, *ad hoc* network applications may require agents to reach a consensus on the average of their measurements [3], including distributed coordination of mobile autonomous agents [4] and distributed data fusion in sensor networks [5]. Another example is the large scale machine learning where a computation task may be executed by collaborative microprocessors with individual memories and storage spaces [6, 7]. Many of the distributed optimization problems, such as those mentioned above, can be cast as an optimization problem in which a network of N agents cooperatively solve

$$\underset{\tilde{x}}{\text{minimize}} \sum_{i=1}^N f_i(\tilde{x}), \quad (1)$$

over a common variable \tilde{x} , where $f_i : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{\infty\}$ is the local objective function associated with agent i . The function f_i is composed of a smooth component $g_i : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{\infty\}$ and a non-smooth component $h_i : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{\infty\}$. Examples of such models include least squares [8, 9] and regularized least squares [10–12]. The variable \tilde{x} may represent average temperature of a room [5], frequency-domain occupancy of spectra [12], states of smart grid systems [13], etc.

This paper studies algorithms that solve (1) in a distributed manner, i.e., agents exchange information only with their immediate neighbors. Such algorithms are extremely attractive for large scale networks that are characterized by the lack of centralized access to

information. They also are energy efficient and enhance the survivability of the network compared with fusion center based processing [8]. Existing distributed approaches include incremental algorithm [14], distributed subgradient descent algorithm [15, 16], dual averaging method [17], and the alternating direction method of multipliers (ADMM) [18–20]. While the above algorithms have been investigated extensively in the literature, we consider a practical constraint, *quantization*, on inter-agent communications. This is largely motivated by a number of physical factors, such as limited bandwidth, sensor battery power, and computing resources, which place tight constraints on the rate and form of information exchange amongst neighboring nodes. Methods that handle this constraint include quantized incremental algorithm [21], quantized dual averaging [22], and quantized subgradient method [16]. These methods are appealing due to its simplicity and ability to handle a wide range of problems. However, they usually proceed slowly to a neighborhood of the optimal solution and the errors are much undesired when the network becomes large.

Recently, the ADMM has been shown to be an efficient algorithm for large scale optimizations and used in various applications such as regression and classification [18]. The authors of [20, 23–25] and [8] have shown its fast convergence rate and resilience to noise, respectively. Considering that quantization operation is equivalent to adding some noise to the data, we expect a quantized distributed ADMM algorithm that works well for Problem (1) in terms of both consensus error and convergence time.

Our main contribution is to develop and analyze a quantized distributed algorithm based on the ADMM. Compared with our previous work [9] where the local objective functions are quadratic, the problem setting and proof of this paper are more general. Indeed, the general local objective functions make the problem much harder because of the possible nonlinearity of their gradients: the effect of dithered quantization (see, e.g., [26]) is not easy to characterize due to this nonlinearity. We thereby adopt deterministic quantization in this paper. We establish that this quantized algorithm converges to a consensus within finite iterations under similar assumptions in [20] as long as an initialization condition is satisfied. The initialization condition is rather mild; simply setting all the variables to 0 suffices. We derive a tight upper bound for the consensus error which does not depend on the size of the network. We finally characterize the convergence time, that is, our algorithm converges within $\log_{1+\eta} \Omega$ iterations, where $\eta > 0$ depends on the network topology and the local objectives, and Ω is a polynomial fraction depending on the quantization resolution, the distance between initial and optimal variable values, the local objectives, and the network topology. Numerical examples validate the fast convergence and small consensus errors compared with existing algorithms.

Notations: We use 0 to denote the all-zero column vector with suitable dimensions. 0_K and I_K are the $K \times K$ all-zero and identity matrix, respectively. Denote $\|x\|_2$ as the Euclidean norm of a vector

x . Given a positive semidefinite matrix G with proper dimensions, the G -norm of x is $\|x\|_G = \sqrt{x^T G x}$. Denote $\sigma_{\max}(D)$ as the largest singular value of a matrix D and $\tilde{\sigma}_{\min}(D)$ as the smallest nonzero singular value of D . $\partial f(x)$ denotes a subgradient of f at x for a convex function $f(x)$ while $\nabla f(x)$ denotes the gradient if it is known to be differentiable.

2. DISTRIBUTED OPTIMIZATION VIA THE ADMM

This section briefly reviews the consensus ADMM (C-ADMM) for multi-agent distributed optimization where agents can send and receive real data with infinite precision. We start with the problem setting and assumptions.

2.1. Problem Setting and Assumptions

Consider a connected network consisting of N agents and E edges, where each agent i has its own objective function $f_i : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{\infty\}$. Assume that the network topology is fixed throughout the paper. We describe this network as a symmetric directed graph $\mathcal{G}_d = \{\mathcal{V}, \mathcal{A}\}$, where \mathcal{V} is the set of vertices with cardinality $|\mathcal{V}| = N$ and \mathcal{A} is the set of arcs with $|\mathcal{A}| = 2E$. Denote M_+ and M_- as the unoriented and oriented incidence matrices with respect to \mathcal{G}_d [27], respectively. Based on this graph, we would like to develop in-network algorithms that find the global optimum \tilde{x}^* (not necessarily unique) minimizing $\sum_{i=1}^N f_i(\tilde{x})$.

We make the following assumptions on the local objective functions $f_i, i = 1, 2, \dots, N$.

Assumption 1 *The local objective functions are proper closed convex functions; for every \tilde{x} where $f_i(\tilde{x})$ is well defined and $f_i(\tilde{x}) < \infty$, there exists at least one bounded subgradient $\partial f_i(\tilde{x})$ such that*

$$f_i(\tilde{y}) \geq f_i(\tilde{x}) + (\partial f_i(\tilde{x}))^T (\tilde{y} - \tilde{x}), \forall \tilde{y} \in \mathbb{R}^M.$$

Moreover, the minimum of (1) can be attained.

Assumption 2 *The smooth components have Lipschitz continuous gradients, i.e., for each agent i there exists a $M_{g_i} > 0$ such that*

$$\|\nabla g_i(\tilde{x}) - \nabla g_i(\tilde{y})\|_2 \leq M_{g_i} \|\tilde{x} - \tilde{y}\|_2, \forall \tilde{x}, \tilde{y} \in \mathbb{R}^M.$$

In addition, the smooth components are strongly convex, i.e., for each agent i there exists a $m_{g_i} > 0$ such that

$$(\nabla g_i(\tilde{x}) - \nabla g_i(\tilde{y}))^T (\tilde{x} - \tilde{y}) \geq m_{g_i} \|\tilde{x} - \tilde{y}\|_2^2, \forall \tilde{x}, \tilde{y} \in \mathbb{R}^M.$$

Assumption 3 *The non-smooth components are convex.*

Note that Assumption 2 implies the differentiability of g_i . Assumptions 1–3 together indicate that Problem (1) has a unique and attainable solution, i.e., $\tilde{x}^* \in \mathbb{R}^M$ is unique.

2.2. The ADMM for Distributed Optimization: C-ADMM

To solve (1) using the ADMM, we first reformulate it as

$$\begin{aligned} & \underset{\{x_i\}, \{z_{ij}\}}{\text{minimize}} && \sum_{i=1}^N f_i(x_i) \\ & \text{subject to} && x_i = z_{ij}, x_j = z_{ij}, \forall (i, j) \in \mathcal{A}, \end{aligned} \quad (2)$$

where $x_i \in \mathbb{R}^M$ is the local copy of the common optimization variable \tilde{x} at agent i and $z_{ij} \in \mathbb{R}^M$ is an auxiliary variable imposing

the consensus constraint on neighboring agents i and j . As the given network is connected, the consensus constraint ensures the consensus to be achieved over the entire network, i.e., $x_i = x_j, \forall (i, j) \in \mathcal{A}$, which in turn guarantees that (2) is equivalent to (1). A distributed optimization algorithm then can be obtained by applying the ADMM update [18] with proper initializations. Specifically, it has been shown in [20] that the ADMM update of (2) yields

$$\begin{aligned} x_i^{k+1} &= (\partial f_i + 2\rho |\mathcal{N}_i| I_M)^{-1} \left(\rho |\mathcal{N}_i| x_i^k + \rho \sum_{j \in \mathcal{N}_i} x_j^k - \alpha_i^k \right), \\ \alpha_i^{k+1} &= \alpha_i^k + \rho \left(|\mathcal{N}_i| x_i^{k+1} - \sum_{j \in \mathcal{N}_i} x_j^{k+1} \right) \end{aligned} \quad (3)$$

at node i , where ρ is a positive algorithm parameter, \mathcal{N}_i is the set of neighbors of node i , and $\alpha_i^k \in \mathbb{R}^M$ is the local Lagrangian multiplier of node i . Obviously, (3) is fully decentralized as the update of x_i^{k+1} and α_i^{k+1} only relies on local and neighboring information. We refer to (3) as the C-ADMM.

While the convergence of the C-ADMM under Assumption 1 follows directly from global convergence of the ADMM [18, 23, 25], we will state in Theorem 1 the linear convergence of the C-ADMM, which is the key to proving Theorems 2 and 3. Let $x \in \mathbb{R}^{NM}$ be the vector concatenating all x_i and $\alpha \in \mathbb{R}^{NM}$ the vector concatenating all α_i . Let $g(x) = \sum_{i=1}^N g_i(x_i)$. Define $1_N \in \mathbb{R}^{NM \times M}$ as the matrix composed of $N \times 1$ blocks of identity matrices I_M . Then we have the following theorem.

Theorem 1 ([20, Theorem 1]) *Consider the C-ADMM iteration (3) that solves (2). If f_i is purely smooth, i.e., $f_i = g_i$, Assumptions 1 and 2 hold, and α^0 is initialized in the column space of $M_- M_-^T$, then α^k lies in the columns space of $M_- M_-^T$ for $k = 0, 1, \dots$, and*

$$x^k \rightarrow 1_N \tilde{x}^* \text{ and } \alpha^k \rightarrow \nabla g(1_N \tilde{x}^*) \text{ } R\text{-linearly,}$$

where \tilde{x}^* is the unique solution to Problem (1).

Note that simply setting $\alpha^0 = 0$ ensures that the initialization condition, i.e., α^0 lies in the column space of $M_- M_-^T$, is met. Without this initialization condition, the convergence of $x^k \rightarrow 1_N \tilde{x}^*$ is still true, but the linear convergence rate is not guaranteed.

3. QUANTIZED CONSENSUS ADMM

To model the effect of quantized communication, we assume that each agent can store and compute real values with infinite precision; an agent, however, can only transmit quantized data through the channel which are received by its neighbors without any error. Given a quantization resolution $\Delta > 0$, define the quantization lattice in \mathbb{R} by

$$\Lambda = \{t\Delta : t \in \mathbb{Z}\}.$$

A quantizer is a function $Q : \mathbb{R} \rightarrow \Lambda$ that maps a real value to some point in Λ . Among all deterministic quantizers, we consider the rounding quantizer that projects $y \in \mathbb{R}$ to its nearest point in Λ :

$$Q(y) = t\Delta, \text{ if } \left(t - \frac{1}{2}\right)\Delta \leq y < \left(t + \frac{1}{2}\right)\Delta.$$

Quantizing a vector means quantizing each of its entries. For $w \in \mathbb{R}^L, L \in \mathbb{Z}^+$, the rounding quantizer projects w to its nearest point in Λ^L ; we use $w_{[Q]}$ to denote the quantizer output of w . If we define $e = w_{[Q]} - w$ as the quantization error, the quantization operation

can be viewed as adding the error to the original data, i.e., $w_{[Q]} = w + e$. Note also that the quantization error is bounded, which is given by $\|e\|_2 \leq \frac{1}{2}\Delta\sqrt{L}$.

Now using the above rounding quantization to modify the C-ADMM to meet the communication constraint, we obtain the quantized consensus ADMM (QC-ADMM) in Algorithm 1.

Algorithm 1 QC-ADMM for multi-agent distributed optimization

Require: Initialize $x_i^0 = 0$ and $\alpha_i^0 = 0$ for each agent $i, i = 1, 2, \dots, N$. Set $\rho > 0$ and $k = 0$.

- 1: **repeat**
- 2: every agent i **do**

$$x_i^{k+1} \leftarrow (\partial f_i + 2\rho|\mathcal{N}_i|I_M)^{-1} \left(\rho|\mathcal{N}_i|x_{i[Q]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q]}^k - \alpha_i^k \right), \quad (4)$$

$$\alpha_i^{k+1} \leftarrow \alpha_i^k + \rho \left(|\mathcal{N}_i|x_{i[Q]}^{k+1} - \sum_{j \in \mathcal{N}_i} x_{j[Q]}^{k+1} \right).$$

- 3: **set** $k = k + 1$.
 - 4: **until** a predefined stopping criterion (e.g., a maximum iteration number) is satisfied.
-

By writing $x_{i[Q]}^k = x_i^k + e_i^k$ where e_i^k denotes the quantization error, we can view the QC-ADMM as an ideal C-ADMM update on $x_{[Q]}^k$ and α^k followed by adding an error term caused by quantization. The rest of this paper is devoted to studying the effect of the rounding quantization on the C-ADMM, i.e., the QC-ADMM iteration (4).

3.1. Convergence Results: Smooth Objective Functions

We first study the simple case where local objective functions only contain smooth components, i.e., $f_i = g_i$. To proceed, we state the following lemma.

Lemma 1 ([9, Lemma 2]) *Given a connected network, if α lies in the column space of $M_-M_-^T$, then there exists a unique β lying in the column space of M_-^T such that $\alpha = M_- \beta$.*

According to Theorem 1, initializing α^0 in the column space of $M_-M_-^T$ ensures that the C-ADMM yields the optimal values $(1_N \tilde{x}^*, \nabla g(1_N \tilde{x}^*))$ with $\nabla g(1_N \tilde{x}^*)$ lying in the column space of $M_-M_-^T$. Therefore, Lemma 1 implies that there exists a unique β^* in the column space of M_-^T such that $\nabla g(1_N \tilde{x}^*) = M_- \beta^*$. Define $u^* = [\frac{1}{2}M_+^T 1_N \tilde{x}^*; \beta^*]$. Our main result is stated as follows.

Theorem 2 *Let $f_i = g_i$. Consider the QC-ADMM in Algorithm 1, and suppose that Assumptions 1 and 2 hold. Then we have*

1. *Convergence: the sequence $(x_{[Q]}^k, \alpha^k)$ generated by (4) converges to a finite value $(1_N \tilde{x}_Q^*, \alpha^*)$ as $k \rightarrow \infty$, where \tilde{x}_Q^* is some vector in Λ^M and not necessarily equal to \tilde{x}^* , and $\alpha^* \in \mathbb{R}^{NM}$.*
2. *Consensus error: the consensus error is bounded by*

$$\|\tilde{x}_Q^* - \tilde{x}^*\|_2 \leq \left(\frac{1}{2} + \rho \frac{2E}{\sum_{i=1}^N m_{g_i}} \right) \sqrt{M}\Delta.$$

3. *Number of iterations: $(x_{[Q]}^k, \alpha^k)$ converges within $\lceil \log_{1+\eta} \Omega \rceil$ iterations, where $\eta = \sqrt{1+\delta} - 1$, $m_g \triangleq \min_i \{m_{g_i}\}$, $M_g \triangleq \max_i \{M_{g_i}\}$,*

$$\delta = \min \left\{ \frac{\tilde{\sigma}_{\min}^2(M_-)}{3\sigma_{\max}^2(M_+)}, \frac{8\rho m_g \tilde{\sigma}_{\min}^2(M_-)}{2\rho^2 \sigma_{\max}^2(M_+) \tilde{\sigma}_{\min}^2(M_-) + 3M_g^2} \right\},$$

$$G = \begin{bmatrix} \rho I_{2EM} & 0_{2EM} \\ 0_{2EM} & \frac{1}{\rho} I_{2EM} \end{bmatrix},$$

$$\tau_0 = \frac{1}{4}\Delta \sqrt{M(\sigma_{\max}(M_+)^2 + \sigma_{\max}(M_-)^2)},$$

$$\Omega = \max \left\{ \frac{3\sqrt{\rho}\sigma_{\max}(M_-)(1+\eta)(\|u^*\|_G + \tau_0)}{\eta\Delta}, \frac{3(1+\eta)(\|u^*\|_G + \tau_0)}{\sqrt{2\rho E}\eta\Delta} \right\},$$

and $\lceil y \rceil, y \in \mathbb{R}$, means the smallest integer that is greater than or equal to y ,

Due to the space limitation, the proof of Theorem 2 is presented in [28]. Our remarks regarding this theorem follow.

Remark 1 *An interesting observation is the parameter ρ . While ρ directly affects the consensus error bound as seen from Theorem 2, it is not easy to characterize how it affects the convergence time. We do not study the optimal selection of ρ here but simply set $\rho = 1$ in the sequel. Therefore we do not regard ρ as a factor affecting our algorithm's performance. We refer readers to [18, 20, 29] for discussions on the effect of ρ on the ADMM.*

Remark 2 *The main result for the rounding quantizer also applies to other deterministic quantizers as our proof only uses the deterministic scheme and the bounded quantization error. Note that $x_{i[Q]}^k$ in the QC-ADMM must be quantized for the $(k+1)$ th update at its own node while the local Lagrangian multiplier α_i^k is not. The reason is to guarantee that $\alpha^k = [\alpha_1^k; \alpha_2^k; \dots; \alpha_N^k]$ lies in the column space of $M_-M_-^T$, and hence the QC-ADMM possesses the linear convergence rate at each C-ADMM update.*

3.2. Convergence Results: General Objective Functions

We now consider the general case where $f_i = g_i + h_i$. Even though the QC-ADMM iterations remain the same as in (4), our previous proof does not directly apply as the linear convergence of the C-ADMM is not preserved. We seek to find a class of objective functions such that similar results of Theorem 2 exist. To this end, we make an additional assumption on the smooth component g_i :

Assumption 4 *The solution to $\min_{\tilde{x}} \sum_{i=1}^N g_i(\tilde{x})$ is attainable.*

This assumption together with Assumption 2 implies that $\min_{\tilde{x}} \sum_{i=1}^N g_i(\tilde{x})$ has a unique solution \tilde{x}^* . We can similarly define $u^* = [\frac{1}{2}M_-^T 1_N \tilde{x}^*; \beta^*]$ where β^* is the unique vector in the column space of M_-^T such that $M_- \beta^* = \nabla g(1_N \tilde{x}^*)$. Let x_i^{k+1} denote the $(k+1)$ th update of the C-ADMM from $x_{[Q]}^k$ and α^k where only g_i 's are involved, i.e.,

$$x_i^{k+1} = (\nabla g_i + 2\rho|\mathcal{N}_i|I_M)^{-1} \left(\rho|\mathcal{N}_i|x_{i[Q]}^k + \rho \sum_{j \in \mathcal{N}_i} x_{j[Q]}^k - \alpha_i^k \right),$$

and denote the vector concatenating x_i^{k+1} as $x^{k+1} \in \mathbb{R}^{NM}$. Notice that $x_{[Q]}^{k+1}$ and α^{k+1} are still updated using the QC-ADMM. Then we obtain the following theorem.

Theorem 3 Consider the QC-ADMM algorithm. Suppose that Assumptions 1–4 hold. If $\|x^{k+1} - x^k\|_2 \leq \Delta_x$ for some $\Delta_x > 0$ throughout the iterations, we have the same results as Theorem 2 where τ_0 is replaced with

$$\tau_1 = \left(\frac{1}{2}\Delta_x + \frac{1}{4}\Delta\sqrt{M} \right) \sqrt{\rho\sigma_{\max}^2(M_+) + \frac{1}{\rho}\sigma_{\max}^2(M_-)}.$$

The idea of proof is exactly the same as Theorem 2 hence omitted. We give two commonly used examples of the non-smooth components that satisfy the condition $\|x^{k+1} - x^k\|_2 \leq \Delta_x$ (see also [28] for the proof). They are:

1. ℓ_1 -norm: Let $\|\cdot\|_1$ denote the ℓ_1 -norm and define $h_i(\tilde{x}) = \xi_i \|\tilde{x}\|_1$ with $\xi_i > 0$. Define $|\mathcal{N}|_{\min} = \min_i \{|\mathcal{N}_i|\}$. Then

$$\|x^{k+1} - x^k\|_2 \leq \frac{\left(M \sum_{i=1}^N \xi_i^2 \right)^{1/2}}{m_g + 2\rho|\mathcal{N}|_{\min}}.$$

2. Indicator function with compact box sets: For a non-empty set \mathcal{X} , define its indicator function as

$$I_{\mathcal{X}}(w) = \begin{cases} 0 & \text{if } w \in \mathcal{X}, \\ \infty & \text{otherwise.} \end{cases}$$

Let $h_i(\tilde{x}) = I_{\mathcal{X}}(\tilde{x})$ where \mathcal{X} is a nonempty compact box set, i.e., $\mathcal{X} = \{\tilde{x} \in \mathbb{R}^M : a \preceq \tilde{x} \preceq b\}$ where $a, b \in \mathbb{R}^M$ and \preceq represents the component-wise inequality. Note that this includes the average consensus problem with bounded quantization [30] as a special case. Let $t_i = \max\{\|\nabla g_i(\tilde{x})\|_2 \mid \tilde{x} \in \mathcal{X}\}$ which exists due to Assumption 2 and $Q_0 = \max\{\|\tilde{x}\|_2 + \frac{1}{2}\Delta\sqrt{M} \mid \tilde{x} \in \mathcal{X}\}$. Then

$$\|x^{k+1} - x^k\|_2 \leq \sum_{i=1}^N \frac{(\sqrt{M}+1)t_i + (6\sqrt{M}+4)\rho|\mathcal{N}_i|Q_0}{m_{g_i} + 2\rho|\mathcal{N}_i|}.$$

4. SIMULATIONS

To construct a connected network with N nodes and E edges, we first generate a complete graph consisting of N nodes, and then uniformly randomly remove $\frac{N(N-1)}{2} - E$ edges while ensuring that the network stays connected.

Consider a constrained distributed optimization problem:

$$\begin{aligned} & \underset{\tilde{x}}{\text{minimize}} && \sum_{i=1}^N |a_i| \|\tilde{x}\|_2^2 + b_i^T \tilde{x} \\ & \text{subject to} && -\mathbf{1}_3 \preceq \frac{\tilde{x}}{N} \preceq \mathbf{1}_3, \end{aligned}$$

where $\tilde{x} \in \mathbb{R}^3$, $\mathbf{1}_3 \in \mathbb{R}^3$ is the all-one vector, $a_i \sim \mathcal{N}(0, 1)$, and $b_i \in \mathbb{R}^3$ whose entries follow $\mathcal{N}(0, N^2)$. We use the QC-ADMM and the quantized dual averaging (Q-DA) method to solve this problem. Set $\Delta = 1$. The parameter ρ is chosen to be 1 in the QC-ADMM and the proximal function is chosen as $e(x) = \frac{1}{2}\|x\|_2^2$ for the Q-DA. Fig. 1 shows the simulation results of the network consisting of $N = 40$ nodes and $E = \{300, 600\}$ edges where the maximum iterative error is defined by $\max_{i=1}^N \|x_{i[Q]}^k - \tilde{x}^*\|_2$.

As seen from Fig. 1, the QC-ADMM has a much smaller maximum iterative error and faster convergence speed than the Q-DA. Note that the QC-ADMM converges to a consensus in finite steps

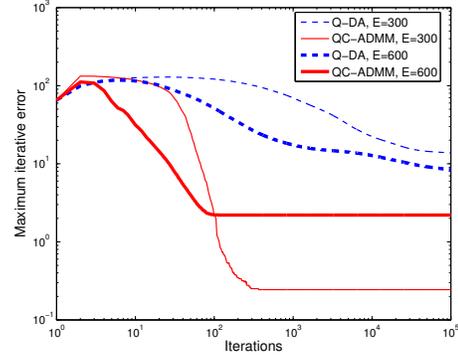


Fig. 1. Performance of the QC-ADMM and the Q-DA where the plotted values are the average of 100 runs.

while the Q-DA does not ensure the convergence nor the consensus (see [22]).

Next, we consider a distributed LASSO problem to study the effect of the quantization resolution on the QC-ADMM:

$$\underset{\tilde{x}}{\text{minimize}} \quad \sum_{i=1}^N \|A_i \tilde{x} - y_i\|_2 + \lambda_i \|\tilde{x}\|_1,$$

where $\tilde{x} \in \mathbb{R}^{20}$ is the unknown variable, $A_i \in \mathbb{R}^{20 \times 20}$ is the linear measurement matrix of agent i whose elements follow $\mathcal{N}(0, 1)$, $y_i \in \mathbb{R}^{20}$ is the measurement vector of agent i whose elements follow $\mathcal{N}(0, N^2)$, and $\lambda_i \in \mathbb{R}^+$ is a positive weight at agent i and follows $\mathcal{N}(0, N^2)$. Set $\rho = 1$ and run the BQ-CADMM with different Δ . Fig. 2 provides the result of a network with $N = 30$ and $E = 200$. The iterative error is defined as $\frac{1}{\sqrt{N}} \|x_{[Q]}^k - \frac{1}{N} \tilde{x}\|_2$ which is equal to the consensus error when a consensus is reached and $\Delta = 0$ corresponds to the ideal case where agents can communicate real values of infinite precision.

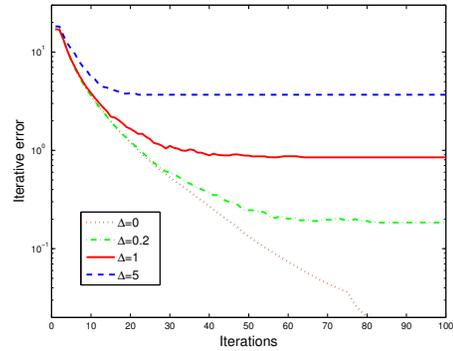


Fig. 2. Solving distributed LASSO using the QC-ADMM.

As we can see from Fig. 2, the consensus error becomes larger as Δ increases. This is not surprising as the higher the quantization resolution is, the more information is lost at each update, thus resulting in a higher consensus error. Meanwhile, the convergence time decreases when Δ increases. This can be seen from the upper bound on number of iterations that guarantees the convergence; that is, Ω decreases as Δ becomes larger. On the other hand, a larger Δ indicates a sparser quantization lattice which makes it easier for the QC-ADMM to reach a convergence point.

5. REFERENCES

- [1] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Nashua, NH, USA: Athena Scientific, 1997.
- [2] G. Giannakis, Q. Ling, G. Mateos, I. Schizas, and H. Zhu, “Decentralized learning for wireless communications and networking,” *ArXiv*, 2015.
- [3] W. Ren, R. W. Beard, and E. M. Atkins, “Information consensus in multivehicle cooperative control,” *IEEE Contr. Syst. Mag.*, vol. 27, no. 2, pp. 71–82, Apr. 2007.
- [4] N. Lynch, *Distributed Algorithms*. San Francisco, CA: Morgan Kaufmann, 1996.
- [5] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” *Proc. Int. Conf. Information Processing in Sensor Networks*, Los Angeles, CA, Apr. 2005.
- [6] R. Bekkerman, M. Bilenko, and J. Langford, *Scaling Up Machine Learning: Parallel and Distributed Approaches*. Cambridge University Press, 2012.
- [7] G. R. Andrews, *Foundations of Multithreaded, Parallel, and Distributed Programming*. Addison-Wesley, 2007.
- [8] H. Zhu, G. B. Giannakis, and A. Cano, “Distributed in-network channel decoding,” *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3970–3983, 2009.
- [9] S. Zhu and B. Chen, “Quantized consensus by the ADMM: probabilistic versus deterministic quantizers,” *IEEE Trans. Signal Process.*, to be published. doi: 10.1109/TSP.2015.2504341.
- [10] G. Mateos, J. Bazerque, and G. Giannakis, “Distributed sparse linear regression,” *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 1856–1871, 2009.
- [11] Q. Ling and Z. Tian, “Decentralized sparse signal recovery for compressive sleeping wireless sensor networks,” *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3816–3827, 2010.
- [12] J. Bazerque and G. Giannakis, “Distributed spectrum sensing for cognitive radio networks by exploiting sparsity,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, 2010.
- [13] L. Gan, U. Topcu, and S. Low, “Optimal decentralized protocol for electric vehicle charging,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 940–951, 2013.
- [14] A. Nedić and D. P. Bertsekas, “Incremental subgradient methods for nondifferentiable optimization,” *SIAM J. Optim.*, vol. 56, no. 1, pp. 109–138, 2001.
- [15] A. Nedić and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [16] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, “Distributed subgradient methods and quantization effects,” in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, 2008.
- [17] J. Duchi, A. Agarwal, and M. Wainwright, “Dual averaging for distributed optimization: convergence analysis and network scaling,” *IEEE Trans. Autom. Control*, vol. 57, no.3, pp. 592–606, 2012.
- [18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundat. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [19] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Puschel, “D-ADMM: A communication-efficient distributed algorithm for separable optimization,” *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 2718–2723, 2013.
- [20] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, “On the linear convergence of the ADMM in decentralized consensus optimization,” *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1760, 2014.
- [21] M. Rabbat and R. Nowark, “Quantized incremental algorithms for distributed optimization,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, 2006.
- [22] D. Yuan, S. Xu, H. Zhao, and L. Rong, “Distributed dual averaging method for multi-agent optimization with quantized communication,” *Syst. Control Lett.*, vol. 61, no. 11, pp. 1043–1061, 2012.
- [23] B. He and X. Yuan, “On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method,” *SIAM J. Numer. Anal.*, vol. 50, no. 2 pp. 700–709, 2012.
- [24] M. Hong and Z.-Q. Luo, “On the linear convergence of the alternating direction method of multipliers,” *ArXiv*, 2013.
- [25] W. Deng and W. Yin, “On the global and linear convergence of the generalized alternating direction method of multipliers,” *Rice Univ. CAAM Tech Report*, TR12-14, 2012.
- [26] L. Schuchman, “Dither signals and their effect on quantization noise,” *IEEE Trans. Commun. Technol.*, vol. COMM-12, pp. 162–165, 1964.
- [27] F. Chung, *Spectral Graph Theory*, Amer. Math. Soc., Providence, 1997.
- [28] S. Zhu, M. Hong, and B. Chen “Quantized consensus ADMM for multi-agent distributed optimization,” *ArXiv*, 2015.
- [29] E. Ghadimi, A. Teixeira, I. Shames and M. Johansson, “Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems,” *IEEE Trans. Autom. Control*, vol. 60, no.3, pp. 644–658, 2015.
- [30] S. Zhu and B. Chen, “Distributed average consensus with bounded quantization,” *ArXiv*, 2016.