Title
Quantized Decentralized Consensus Optimization

Permalink
https://escholarship.org/uc/item/35n3h6f2

Authors
Reisizadeh, Amirhossein
Mokhtari, Aryan
Hassani, Hamed
et al.

Publication Date
2019-01-21

Peer reviewed
Quantized Decentralized Consensus Optimization

Amirhossein Reisizadeh, Aryan Mokhtari, Hamed Hassani, and Ramtin Pedarsani

Abstract—We consider the problem of decentralized consen-
sus optimization, where the sum of $n$ convex functions are
minimized over $n$ distributed agents that form a connected net-
work. In particular, we consider the case that the communicated
local decision variables among nodes are quantized in order to
alleviate the communication bottleneck in distributed optimiza-
tion. We propose the Quantized Decentralized Gradient Descent
(QDGD) algorithm, in which nodes update their local decision
variables by combining the quantized information received from
their neighbors with their local information. We prove that un-
der standard strong convexity and smoothness assumptions for
local cost functions, QDGD achieves a vanishing mean solution
error. To the best of our knowledge, this is the first algorithm
that achieves vanishing consensus error in the presence of quantization noise. Moreover, we provide simulation results
that show tight agreement between our derived theoretical
convergence rate and the experimental results.

I. INTRODUCTION

Distributed optimization of a sum of convex functions has
a variety of applications in different areas including de-
centralized control systems [1], wireless systems [2], sensor
networks [3], networked multiagent systems [4], multirobot
networks [5], and large scale machine learning [6]. In such
problems, one aims to solve a consensus optimization prob-
lem to minimize $f(x) = \sum_{i=1}^{n} f_i(x)$ cooperatively over $n$
nodes or agents that form a connected network. The function
$f_i(\cdot)$ represents the local cost function of node $i$ that is only
known by this node.

Distributed optimization has been largely studied in the
literature starting from seminal works in the 80s [7], [8].
Since then, various algorithms have been proposed to ad-
dress decentralized consensus optimization in multiagent
systems. The most commonly used algorithms are decen-
tralized gradient descent or gradient projection method [9–
[12], distributed alternating direction method of multipliers
(ADMM) [13]–[15], decentralized dual averaging [16], [17],
and distributed Newton optimization method [18], [19]. Fur-
thermore, the decentralized consensus optimization problem
has been considered in online or dynamic settings, where
the dynamic cost function becomes an online regret function
[20], [21].

A major bottleneck in achieving fast convergence in de-
centralized consensus optimization is limited communication
bandwidth among nodes. As the dimension of input data
increases (which is the current trend in large-scale distributed
machine learning), a considerable amount of information
must be exchanged among nodes, over many iterations of
the consensus algorithm. This causes a significant commu-
nication bottleneck that can substantially slow down the
corvergence time of the algorithm [22], [23].

Quantized communication for the agents is brought into
the picture for bounded and stable control systems [24].
Furthermore, consensus distributed averaging algorithms are
studied under discretized message passing [25]. Motivated
by the energy and bandwidth-constrained wireless sensor
networks, the work in [26] proposes distributed optimization
algorithms under quantized variables and guarantees conver-
genence within a non-vanishing error. Deterministic quantiza-
tion has been considered in distributed averaging algorithms
[27] where the iterations converge to a neighborhood of
the average of initials. However, randomized quantization
schemes are shown to achieve the average of initials, in
expectation [28]. The work in [29] also considers a con-
sensus distributed optimization problem over a cooperative
network of agents restricted to quantized communication.
The proposed algorithm guarantees convergence to the op-
tima within an error which depends on the network size
and the number of quantization levels. Aligned with the
communication bottleneck described earlier, [30] provides a
quantized distributed load balancing scheme that converges
to a set of desired states while the nodes are constrained to
remain under maximum load capacities.

More recently, 1–Bit SGD [22] was introduced in which
at each time step, the agents sequentially quantize their local
gradient vectors by entry-wise signs while contributing the
quantization error induced in previous iteration. Moreover,
in [31], the authors propose the Quantized-SGD (QSGD), a
class of compression scheme algorithms that is based on a
stochastic and unbiased quantizer of the vector to be trans-
mittted. QSGD provably provides convergence guarantees, as
well a good practical performance. Recently, a different line
of work has proposed the use of coding theoretic techniques
to alleviate the communication bottleneck in distributed
computation [32]–[35].

In this paper, our goal is to analyze the quantized de-
centralized consensus optimization problem, where node $i$
transmits a quantized version of its local decision vari-
able $Q(x_i)$ to the neighboring nodes instead of the exact
decision variable $x_i$. Motivated by the stochastic quanti-
tizer proposed in [31], we consider unbiased and variance
bounded random quantizers \(Q(\cdot)\), i.e. \(\mathbb{E}[Q(x)|x] = x\) and \(\mathbb{E}[(Q(x) - x)^2|x] \leq \sigma^2\) for some fixed constant \(\sigma^2\).

Our main contribution is to propose a Quantized Decentralized Gradient Descent (QDGD) method, which involves a novel way of updating the local decision variables by combining the quantized message received from the neighbors and the local information such that proper averaging is performed over the local decision variable and the neighbors’ quantized vectors. We prove that under standard strong convexity and smoothness assumptions, for any unbiased and variance bounded quantizer, QDGD achieves a vanishing mean solution error: for all \(i = 1, \ldots, n\) we obtain that \(\mathbb{E}[(x_{i,t} - x^*)^2] \leq O\left(\frac{1}{t^{1/4}}\right)\) for sufficiently large \(T\), where \(x_{i,t}\) is the local decision variable of node \(i\) at iteration \(T\) and \(x^*\) is the optimal solution. To the best of our knowledge, this is the first decentralized gradient-based algorithm that achieves vanishing consensus error in the presence of non-vanishing quantization noise. We further provide simulation results that corroborate our theoretical results.

Notation. In this paper, we denote by \([n]\) the set \(\{1, \ldots, n\}\) for any natural number \(n \in \mathbb{N}\). The gradient of a function \(f(x)\) is denoted by \(\nabla f(x)\). For non-negative functions \(g\) and \(h\) of \(t\), we denote \(g(t) = O(h(t))\) if there exist \(t_0 \in \mathbb{N}\) and constant \(c > 0\) such that \(g(t) \leq c h(t)\) for all \(t \geq t_0\). We use \([a]\) to indicate the least integer greater than or equal to \(a\).

Paper Organization. The rest of the paper is organized as follows. In Section II, we precisely formulate the quantized decentralized consensus optimization problem. We provide the description of the Quantized Decentralized Gradient Descent algorithm in Section III. The main theorem of the paper is stated and proved in Section IV. We provide numerical studies in Section V. Finally, we conclude the paper and discuss future directions in Section VI.

II. PROBLEM FORMULATION

In this section we formally define the consensus optimization problem that we aim to solve. Consider a set of \(n\) nodes that communicate over a connected and undirected graph \(G = (\mathcal{V}, \mathcal{E})\) where \(\mathcal{V} = \{1, \ldots, n\}\) and \(\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}\) denote the set of nodes and edges, respectively. We assume that nodes are only allowed to exchange information with their neighbors and use the notation \(\mathcal{N}_i\) for the set of node \(i\)'s neighbors. In our setting, we assume that each node \(i\) has access to a local convex function \(f_i : \mathbb{R}^p \rightarrow \mathbb{R}\), and nodes in the network cooperate to minimize the aggregate objective function \(f : \mathbb{R}^p \rightarrow \mathbb{R}\) taking values \(f(x) = \sum_{i=1}^n f_i(x)\). In other words, nodes aim to solve the optimization problem

\[
\min_{x \in \mathbb{R}^p} f(x) = \min_{x \in \mathbb{R}^p} \sum_{i=1}^n f_i(x). \tag{1}
\]

We assume the local objective functions \(f_i\) are strongly convex and smooth, and, therefore, the aggregate function \(f\) is also strongly convex and smooth. In the rest of the paper, we use \(x^*\) to denote the unique minimizer of Problem (1).

In decentralized settings, nodes have access to a single summand of the global objective function \(f\) and to reach the optimal solution \(x^*\) communication with neighboring nodes is inevitable. To be more precise, nodes need to minimize their local objective functions, while they ensure that their local decision variables are equal to their neighbors’. This interpretation leads to an equivalent formulation of Problem (1). If we define \(x_i\) as the decision variable of node \(i\), the alternative formulation of Problem (1) can be written as

\[
\begin{align*}
\min_{x_1, \ldots, x_n \in \mathbb{R}^p} & \sum_{i=1}^n f_i(x_i) \\
\text{s.t.} & \ x_i = x_j, \quad \text{for all } i, j \in \mathcal{N}_i. \tag{2}
\end{align*}
\]

Since we assume that the underlying network is a connected graph, the constraint in (2) implies that any feasible solution should satisfy \(x_1 = \cdots = x_n\). Under this condition the objective function values in (1) and (2) are equivalent. Hence, it follows that the optimal solutions of Problem (2) are equal to the optimal solution of Problem (1), i.e., if we denote \(\{x_i^*\}_{i=1}^n\) as the optimal solutions of Problem (2) it holds that \(x_1^* = \cdots = x_n^* = x^*\). Therefore, we proceed to solve Problem (2) which is naturally formulated for decentralized optimization in lieu of Problem (1).

The problem formulation in (2) suggests that each node \(i\) should minimize its local objective function \(f_i\) while keeping its decision variable \(x_i\) close to the decision variable \(x_j\) of its neighbors \(j \in \mathcal{N}_i\). This goal can be achieved by exchanging local variables \(x_i\) among neighboring nodes to enforce consensus on the decision variables. Indeed, exchange of updated local vectors between the distributed nodes induces a potentially heavy communication load on the shared bus. To address this issue, we assume that each node provides a randomly quantized variant of its local updated variable to the neighboring nodes. That is, if we denote by \(x_i\) the decision variable of node \(i\), then the corresponding quantized variant \(z_i = Q(x_i)\) is communicated to the neighboring nodes, \(\mathcal{N}_i\). Exchanging quantized vectors \(z_i\) instead of the true vectors \(x_i\) indeed reduces the communication burden at the cost of injecting noise to the information received by the nodes in the network. The main challenge in this setting is to ensure that nodes can still converge to the optimal solution of Problem (2), while they only have access to a quantized variant of their neighbors’ true decision variables.

III. QDGD ALGORITHM

In this section, we propose a quantized gradient based method to solve the decentralized optimization problem in (2) and consequently the original problem in (1) in a fully decentralized fashion. To do so, consider \(x_{i,t}\) as the decision variable of node \(i\) at step \(t\) and \(z_{i,t} = Q(x_{i,t})\) as the quantized version of the vector \(x_{i,t}\). In the proposed Quantized Decentralized Gradient Descent (QDGD) method, nodes update their local decision variables by combining the quantized information received from their neighbors with their local information. To formally state the update of QDGD, we first define \(w_{ij}\) as the weight that node \(i\) assigns.
Algorithm 1 QDGD at node $i$

Require: Weights $\{w_{ij}\}_{j=1}^{n}$, total iterations $T$
1: Set $x_{i,0} = 0$
2: for $t = 0, \cdots , T - 1$
3: \hspace{1em} $z_{i,t} = Q(x_{i,t})$ to $j \in N_{i}$ and receive $z_{j,t}$
4: Compute $x_{i,t+1}$ according to the update in (3)
5: end for
6: return $x_{i,T}$

to node $j$. If nodes $i$ and $j$ are not neighbors then $w_{ij} = 0$, and if they are neighbors the weight $w_{ij} \geq 0$ is nonnegative. At each time step $t$, each node $i$ sends its quantized $z_{i,t}$ variant of its local vector $x_{i,t}$ to its neighbors $j \in N_{i}$ and receives their corresponding vectors $z_{j,t}$. Then, using the received information it updates its local decision variable according to the update

$$x_{i,t+1} = (1 - \varepsilon + \varepsilon w_{ii})x_{i,t} + \varepsilon \sum_{j \in N_{i}} w_{ij}z_{j,t} - \alpha \varepsilon \nabla f_{i}(x_{i,t}), \quad (3)$$

where $\varepsilon$ and $\alpha$ are positive step-sizes. The update of QDGD in (3) shows that the updated decision variable $x_{i,t+1}$ is evaluated by proper averaging over the local decision variable $x_{i,t}$ and neighbors quantized vectors $z_{j,t}$, and descending through the negative local gradient $\nabla f_{i}(x_{i,t})$ with a proper stepsize. Note that quantized decision variables of the neighboring nodes contribute to the descent direction proportionally to step-size $\varepsilon$, unlike the noiseless local gradient which is scaled by $\alpha \varepsilon$. The steps of the proposed QDGD algorithm are summarized in Algorithm 1.

Remark 1. The proposed QDGD algorithm can be interpreted as a variant of the decentralized (sub)gradient descent (DGD) method [9], [10] for quantized decentralized optimization (see Section IV). Note that the vanilla DGD method converges to a neighborhood of the optimal solution in the presence of quantization noise where the radius of convergence depends on the variance of quantization error [9], [10], [26], [29]. QDGD improves the inexact convergence of DGD by modifying the contribution of quantized information received from neighboring noise as described in update (3).

In particular, as we show in Theorem 1, the sequence of iterates generated by QDGD converges to the optimal solution of Problem (1) in expectation.

Moreover, the proposed QDGD algorithm does not restrict the quantizer, except for few customary conditions. However, design of efficient quantizers has been taken into consideration. For instance, consider a low-precision representation specified by $\gamma \in \mathbb{R}$ and $b \in \mathbb{N}$. The range representable by scale factor $\gamma$ and $b$ bits is $\{-\gamma \cdot 2^{k-1}, \cdots , -\gamma , 0, \gamma, \cdots , \gamma, 2^{b-1} - 1\}$. For any $k \gamma \leq x < (k + 1)\gamma$ in the representable range, the low-precision quantizer outputs

$$Q(\gamma, b)(x) = \begin{cases} k\gamma + \frac{x - k\gamma}{\gamma} & \text{w.p. } 1 - \frac{x - k\gamma}{\gamma}, \\ (k + 1)\gamma & \text{w.p. } \frac{x - k\gamma}{\gamma}. \end{cases} \quad (4)$$

For any $x$ in the range, the quantizer is unbiased and variance bounded, i.e., $E[Q(\gamma, b)(x)] = x$ and $E[\left\|Q(\gamma, b)(x) - x\right\|^2] \leq \frac{\gamma^2}{4}$.

IV. CONVERGENCE ANALYSIS

In this section, we prove that for sufficiently large number of iterations, the sequence of local iterates generated by QDGD converges to an arbitrarily precise approximation of the optimal solution of Problem (1). The following assumptions hold throughout the analysis of the algorithm.

Assumption 1. Local objective functions $f_i$ are differentiable and smooth with parameter $L$, i.e.,

$$\left\|\nabla f_i(x) - \nabla f_i(y)\right\| \leq L\|x - y\|, \quad (5)$$

for any $x, y \in \mathbb{R}^p$.

Assumption 2. Local objective functions $f_i$ are strongly convex with parameter $\mu$, i.e.,

$$\langle \nabla f_i(x) - \nabla f_i(y), x - y \rangle \geq \mu \|x - y\|^2, \quad (6)$$

for any $x, y \in \mathbb{R}^p$.

Assumption 3. The random quantizer $Q(\cdot)$ is unbiased and has a bounded variance, i.e.,

$$E[Q(x)|x] = x, \quad \text{and} \quad E[\left\|Q(x) - x\right\|^2| x] \leq \sigma^2, \quad (7)$$

for any $x \in \mathbb{R}^p$; and quantizations are carried out independently on distributed nodes.

Assumption 4. The weight matrix $W \in \mathbb{R}^{n \times n}$ with entries $w_{ij}$ satisfies the following conditions

$$W = W^T, \quad W1 = 1, \quad \text{null}(I - W) = \text{span}(1). \quad (8)$$

The conditions in Assumptions 1 and 2 imply that the global objective function $f$ is strongly convex with parameter $\mu$ and its gradients are Lipschitz continuous with constant $L$. Assumption 3 poses two customary conditions on the quantizer, that are unbiasedness and variance boundedness. Assumption 4 implies that weight matrix $W$ is symmetric and doubly stochastic. The largest eigenvalue of $W$ is $\lambda_1(W) = 1$ and all the eigenvalues belong to $(-1, 1]$, i.e. the ordered sequence of eigenvalues of $W$ are $1 = \lambda_1(W) \geq \lambda_2(W) \geq \cdots \geq \lambda_n(W) > -1$. We denote by $1 - \beta$ the spectral gap associated to the stochastic matrix $W$, where $\beta = \max\{[\lambda_2(W)], [\lambda_n(W)]\}$ is the second largest magnitude of the eigenvalues of matrix $W$. It is also customary to assume rank$(I - W) = n - 1$ such that null$(I - W) = \text{span}(1)$.

In the following theorem we show that the local iterations generated by QDGD converge to the global optima, as close as desired.

Theorem 1. Consider the distributed consensus optimization Problem (1) and suppose Assumptions 1– 4 hold. Then, for each node $i$, the expected deviation of the output of
Algorithm 1 from the solution to Problem (1) is upper bounded by
\[
E \left[ \| x_{i,T} - x^* \|^2 \right] \leq O \left( \frac{1}{T^{1/4}} \right),
\]
for \( \varepsilon = \frac{c_1}{T^{1/4}}, \alpha = \frac{c_2}{T^{1/4}} \) and \( T \geq T_0 \), where \( c_1, c_2 \) and \( T_0 \) are positive constants independent of \( T \).

**Remark 2.** Theorem 1 demonstrates that the proposed QGDQ provides an approximation solution with vanishing deviation from the optimal solution, despite the fact that the quantization noise does not vanish by iteration.

To analyze the proposed method, we start by rewriting the update rule (3) as follows
\[
x_{i,t+1} = x_{i,t} - \varepsilon_t \left( (1 - w_{ii})x_{i,t} - \sum_{j \neq i} w_{ij}x_{j,t} + \alpha \nabla f_i(x_{i,t}) \right).
\]
(10)

The next step is to write the update (10) in a matrix form. To do so, we define the function \( F : \mathbb{R}^{np} \rightarrow \mathbb{R} \) as \( F(x) = \sum_{i=1}^n f_i(x_i) \) where \( x_i \in \mathbb{R}^p \) and \( x = [x_1; \ldots; x_n] \in \mathbb{R}^{np} \). It is easy to verify that the gradient of the function \( F \) is the concatenation of local gradients evaluated at the local variable, that is \( \nabla F(x_i) = [\nabla f_1(x_{1,t}); \ldots; \nabla f_n(x_{n,t})] \). We also define the matrix \( W = W_D \odot I \in \mathbb{R}^{np \times np} \) as the Kronecker product of the weight matrix \( W \in \mathbb{R}^{n \times n} \) and the identity matrix \( I \in \mathbb{R}^{p \times p} \). Similarly, define \( W_D = W_D \odot I \in \mathbb{R}^{np \times np} \), where \( W_D = [w_{ii}] \in \mathbb{R}^{n \times n} \) denotes the diagonal matrix of the entries on the main diagonal of \( W \). For the sake of consistency, we denote by the boldface \( I \) the identity matrix of size \( np \). According to above definitions, we can write the concatenated version of (10) as follows,
\[
x_{t+1} = x_t - \varepsilon_t \left( (W_D - W)z_t + (I - W_D)x_t + \alpha \nabla F(x_t) \right).
\]
(11)

As we discussed in Section II, the distributed consensus optimization Problem (1) can be equivalently written as Problem (2). The constraint in the latter restricts the feasible set to the consensus vectors, that is \( \{ x = [x_1; \ldots; x_n] : x_1 = \cdots = x_n \} \). According to the discussion on rank of the weight matrix \( W \), the null space of the matrix \( I - W \) is \( \text{null}(I - W) = \text{span}(1) \). Hence, the null space of \( I - W \) is the set of all consensus vectors, i.e. \( x \in \mathbb{R}^{np} \) is feasible for problem (2) if and only if \( (I - W)x = 0 \), or equivalently \( (I - W)^{1/2}x = 0 \). Therefore, the alternative Problem (2) can be compactly represented as the following linearly-constrained problem,
\[
\min_{x \in \mathbb{R}^{np}} F(x) = \sum_{i=1}^n f_i(x_i)
\]
subject to \( (I - W)^{1/2}x = 0 \).
(12)

We denote by \( \bar{x} = [x^*; \ldots; x^*] \) the unique solution to (12).

Now, for given penalty parameter \( \alpha > 0 \), one can define the quadratic penalty function corresponding to the linearly constraint problem (12) as follows,
\[
h_\alpha(x) = \frac{1}{2} x^T (I - W)x + \alpha F(x).
\]
(13)

Since \( I - W \) is a positive semi-definite matrix and \( F \) is \( L \)-smooth and \( \mu \)-strongly convex, therefore the function \( h_\alpha \) is \( L_\alpha \)-smooth and \( \mu_\alpha \)-strongly convex on \( \mathbb{R}^{np} \) having \( L_\alpha = 1 - \lambda_n(W) + \alpha L \) and \( \mu_\alpha = \alpha \mu \). We denote by \( x_\alpha \) the unique minimizer of \( h_\alpha(x) \), i.e.
\[
x_\alpha = \arg \min_{x \in \mathbb{R}^{np}} h_\alpha(x) = \arg \min_{x \in \mathbb{R}^{np}} \frac{1}{2} x^T (I - W)x + \alpha F(x).
\]
(14)

In the following, we link the solution of problem (14) to the local variable iterations provided by Algorithm 1. Specifically, for sufficiently large number of iterations \( T \), we demonstrate that for proper choice of step-sizes, the expected squared deviation of \( x_T \) from \( x_\alpha \) vanishes sub-linearly.

**Lemma 1.** Consider the optimization Problem (14) and suppose Assumptions 1-4 hold. Then, the expected deviation of the output of QGDQ from the solution to Problem (14) is upper bounded by
\[
E \left[ \| x_T - x_\alpha \|^2 \right] \leq O \left( \frac{c_3 n \sigma^2 \| W - W_D \|_F}{\mu c_2} \frac{1}{T^{1/4}} \right),
\]
(15)
for \( \varepsilon = \frac{c_1}{T^{1/4}}, \alpha = \frac{c_2}{T^{1/4}} \) and \( T \geq T_1 \), where \( c_1 \) and \( c_2 \) are positive constants independent of \( T \), and \( T_1 \) is the maximum iterations as stated in Lemmas 1.

**Lemma 1** guarantees convergence of the proposed iterations (3) to the solution of the later-defined Problem (14). Loosely speaking, Lemma 1 ensures that \( x_T \) is close to \( x_\alpha \) for large \( T \). So, in order to capture the deviation of \( x_T \) from the global optima \( x^* \), it suffices to show that \( x_\alpha \) is close to \( x^* \), as well. The following lemma guarantees such argument.

**Lemma 2.** Consider the distributed consensus optimization Problem (1) and the problem defined in (14). If Assumptions 1, 2 and 4 hold, then the deviation of the two solutions is bounded as
\[
\| x_\alpha - \bar{x} \| \leq O \left( \frac{c_2}{1 - \beta} \frac{1}{T^{1/4}} \right),
\]
(16)
for \( \alpha = \frac{c_2}{T^{1/4}} \) and \( T \geq T_2 \), where \( c_2 \) is a positive constant independent of \( T \) and \( T_2 := \max \left\{ \left[ \frac{c_2 L}{1 - \lambda_n(W)} \right]^4, \left[ \frac{c_2^2 (\mu + L)}{c_2^4} \right]^4 \right\} \).

Proofs of Lemmas 1 and 2 are skipped to the long version of the paper [36]. Having set the main lemmas, now it is straightforward to prove Theorem 1.

**Proof of Theorem 1.** For the specified step-sizes \( \varepsilon \) and \( \alpha \) and large enough iterations \( T \geq T_0 := \max\{T_1, T_2\} \), Lemmas 1...
and 2 are applicable and we have

$$\mathbb{E} \left[ \|x_T - \tilde{x}^*\|^2 \right] = \mathbb{E} \left[ \|x_T - x^*_\alpha + x^*_\alpha - \tilde{x}^*\|^2 \right]$$

$$\leq 2\mathbb{E} \left[ \|x_T - x^*_\alpha\|^2 \right] + 2\|x^*_\alpha - \tilde{x}^*\|^2$$

$$\leq O \left( \frac{1}{T^{1/4}} \right) + O \left( \frac{1}{T^{1/2}} \right)$$

$$= O \left( \frac{1}{T^{1/4}} \right).$$  \hfill (17)

Since $\mathbb{E} \left[ \|x_{i,T} - x^*\|^2 \right] \leq \mathbb{E} \left[ \|x_T - \tilde{x}^*\|^2 \right]$ for any $i = 1, \ldots, n$, the inequality in (17) follows the claim of Theorem 1.

\hfill \square

V. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the proposed QDGD Algorithm on minimizing a distributed quadratic objective. We pictorially demonstrate the effect of quantization noise and graph topology on the relative expected error rate.

Consider the quadratic optimization problem

$$\min_{x \in \mathbb{R}^p} f(x) = \sum_{i=1}^{n} \frac{1}{2} x^T A_i x + b_i^T x,$$  \hfill (18)

where $f_i(x) = \frac{1}{2} x^T A_i x + b_i^T x$ denotes the local objective function of node $i \in [n]$. The unique solution to (18) is therefore $x^* = -\left(\sum_{i=1}^{n} A_i\right)^{-1} \left(\sum_{i=1}^{n} b_i\right)$. We pick diagonal matrices $A_i$ such that $p/2$ of the diagonal entries of each $A_i$ are drawn from the set $\{1, 2, 2^2\}$ and the other $p/2$ diagonal entries are drawn from the set $\{1, 2^{-1}, 2^{-2}\}$, all uniformly at random. Entries of vectors $b_i$ are randomly picked from the interval $(0, 1)$. In our setup, the graph of agents is a connected Erdős-Rényi edge weight matrix with edge probability $p_c$. We set the edge weight matrix to be $W = I - \frac{2}{\lambda_{\text{max}}(L)} L$ where $L$ is the Laplacian with $\lambda_{\text{max}}(L)$ as its largest eigenvalue. In our simulations, we let an additive noise model the quantization error, i.e. $Q(x) = x + \eta$ where $\eta \sim N(0, \frac{\sigma^2}{p} I_p)$.

We first consider a connected Erdős-Rényi graph of $n = 50$ nodes and connectivity probability of $p_c = 0.35$ and dimension $p = 20$. Fig. 1 shows the convergence rate corresponding to three values of quantization noise $\sigma^2 \in \{2, 20, 200\}$, compared to the theoretical upper bound derived in Theorem 1 in the logarithmic scale. As expected, Fig. 1 shows that the error rate linearly scales with the quantization noise; however, it does not saturate around a non-vanishing residual, regardless the variance. Moreover, Fig. 1 demonstrates that the convergence rate closely follows the upper bound derived in Theorem 1. For instance, for the plot corresponding to $\sigma^2 = 20$, the relative errors are evaluated as $e_{T_1}/e_0 = 0.0121$ and $e_{T_3}/e_0 = 0.0082$ for $T_1 = 3200$ and $T_2 = 12800$, respectively. Therefore, $e_{T_2}/e_{T_1} \approx 0.68$ which is upper bounded by $(\frac{T_2}{T_1})^{1/4} \approx 0.7$.

To observe the effect of graph topology, quantization noise variance is fixed to $\sigma^2 = 2$ and we varied the connectivity ratio by picking three different values, i.e. $p_c \in \{0.1, 0.5, 1\}$ where $p_c = 1$ corresponds to the complete graph case. As

Fig. 2 depicts, for the same number of iterations, deviation from the optimal solution tends to increase as the graph is gets sparse. In other words, even noisy information of the neighbor nodes improves the gradient estimate for local nodes. It also highlights the fact that regardless of the sparsity of the graph, the proposed QDGD algorithm guarantees the consensus to the optimal solution for each local node, as long as the graph is connected.

VI. CONCLUSION

We proposed the QDGD algorithm to tackle the problem of quantized decentralized consensus optimization. The algorithm updates the local decision variables by combining the quantized messages received from the neighbors and the local information such that proper averaging is performed over the local decision variable and the neighbors’ quantized vectors. We proved that the QGDG algorithm achieves a vanishing consensus error in mean-squared sense, and verified our theoretical results with numerical studies.

An interesting future direction is to establish a fundamental trade-off between the convergence rate of quantized consensus algorithms and the communication. More precisely, given a target convergence rate, what is the minimum number of bits that one should communicate in decentralized consensus? Another interesting line of research is to develop
novel source coding (quantization) schemes that have low computation complexity and are information theoretically near-optimal in the sense that they have small communication load and fast convergence rate.

References