Energy-Efficient Distributed Learning With Coarsely Quantized Signals

Alireza Danaee, Rodrigo C. de Lamare, and Vítor H. Nascimento

***Abstract*—In this work, we present an energy-efficient distributed learning framework using low-resolution ADCs and coarsely quantized signals for Internet of Things (IoT) networks. In particular, we develop a distributed quantization-aware least-mean square (DQA-LMS) algorithm that can learn parameters in an energy-efficient fashion using signals quantized with few bits while requiring a low computational cost. We also carry out a statistical analysis of the proposed DQA-LMS algorithm that includes a stability condition. Simulations assess the DQA-LMS algorithm against existing techniques for a distributed parameter estimation task where IoT devices operate in a peer-to-peer mode and demonstrate the effectiveness of the DQA-LMS algorithm.**

***Index Terms*—Distributed learning, energy-efficient signal processing, adaptive algorithms, coarse quantization.**

I. Introduction

DISTRIBUTED signal processing algorithms are of great relevance for statistical inference in wireless networks and applications such as wireless sensor networks (WSNs) [1] and the Internet of Things (IoT) [2]. These techniques deal with the extraction of information from data collected at nodes that are distributed over a geographic area. Prior work on distributed approaches has studied protocols for exchanging information [3]–[5], adaptive learning algorithms [6], [7], the exploitation of sparse measurements [8], [9], topology adaptation [10], compensation methods for highly correlated input signals [11], and robust techniques against interference and noise [12]. Although there are many studies on the need for data exchange and signaling among nodes as well as their complexity, prior work on energy-efficient techniques is rather limited.

In this context, energy-efficient signal processing techniques have gained a great deal of interest in the last decade or so due to their ability to save energy and promote sustainable development of electronic systems and devices. Electronic devices often exhibit a power consumption that is dependent on the communication module [13], [14] and from a circuit perspective on analog-to-digital converters (ADCs) and decoders [15]. Reducing the number of bits used to represent digital samples can greatly decrease the energy consumption by ADCs [16]. This is key to devices that are battery operated and wireless networks that must keep the power consumption to a low level for sustainability reasons. In particular, prior work on energy efficiency has reported many contributions in signal processing for communications and electronic systems that operate with coarsely quantized signals [17]–[22].

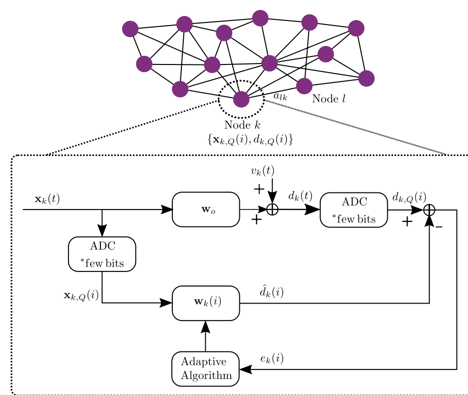


Fig. 1. A distributed adaptive IoT network.

In this work, we propose an energy-efficient distributed learning framework using low-resolution ADCs and coarsely quantized signals for IoT networks. In particular, we devise a distributed quantization-aware least-mean square (DQA-LMS) algorithm that can learn parameters in an energy-efficient way using signals quantized using few bits with a low computational cost. We also develop a statistical analysis of the DQA-LMS algorithm that includes a stability condition. Simulations assess the DQA-LMS algorithm against existing techniques for a distributed parameter estimation task with IoT devices.

This letter is structured as follows: Section II introduces the signal model and states the problem. Section III details the proposed DQA-LMS algorithm, whereas Section IV analyzes DQA-LMS. Section V shows and discusses the simulation results and Section VI draws the conclusions of this work.

II. Signal Model and Problem Statement

We consider an IoT network consisting of N nodes or agents, which run distributed signal processing techniques to perform the desired tasks, as depicted in Fig. 1. The model adopted considers a desired signal $d_k(i)$, at each time i, described by

$$d_k(i) = \mathbf{w}_o^H \mathbf{x}_k(i) + v_k(i), \quad k = 1, 2, \ldots, N, \tag{1}$$

where $\mathbf{w}_o \in \mathbb{C}^{M\times 1}$ is the parameter vector that the agents must estimate, $\mathbf{x}_k(i) \in \mathbb{C}^{M\times 1}$ is the regressor and $v_k(i)$ represents Gaussian noise with zero mean and variance $\sigma^2_{v,k}$ at node k. We adopt the Adapt-then-Combine (ATC) diffusion rule as it outperforms the incremental and consensus protocols [3], [4]. At each node k and time i, based on the local data $\{d_k(i), \mathbf{x}_k(i)\}$ and the estimated parameter vectors $\mathbf{h}_l(i)$ from its neighborhood, the parameter vector with local estimates $\mathbf{w}_k(i)$ is updated. The ATC distributed LMS

Manuscript received November 2, 2020; revised December 23, 2020; accepted January 10, 2021. Date of publication January 13, 2021; date of current version February 11, 2021. This work was supported in part by CNPq, CAPES, FAPERJ and by the ELIOT Project (FAPESP 2018/12579-7, ANR-18-CE40-0030). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arash Mohammadi. *(Corresponding author: Alireza Danaee.)*

Alireza Danaee and Rodrigo C. de Lamare are with the CETUC, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22451-900, Brazil (e-mail: alireza@cetuc.puc-rio.br; rodrigo.delamare@york.ac.uk).

Vítor H. Nascimento is with the Escola Politécnica, University of São, Sao Paulo 05508-970, Brazil (e-mail: vitor@ieee.org).

Digital Object Identifier 10.1109/LSP.2021.3051522

(DLMS) algorithm consists of the recursions:

$$\mathbf{h}_k(i) = \mathbf{w}_k(i-1) + \mu_k \mathbf{x}_k(i) e_k^*(i), \quad \mathbf{w}_k(i) = \sum_{l \in \mathcal{N}_k} a_{lk} \mathbf{h}_l(i),$$

where $\mathbf{h}_k(i)$ and $\mathbf{w}_k(i)$ contain the intermediate and the local estimates of $\mathbf{w}_o$ at node k and time i, respectively, $e_k(i) = d_k(i) - \hat{d}_k(i) = d_k(i) - \mathbf{w}_k^H(i-1)\mathbf{x}_k(i)$ is the error between the output of the adaptive filter, $\hat{d}_k(i)$, and the desired signal, $d_k(i)$, at time i, μ_k is the step-size for node k, $\mathcal{N}_k$ is the set of neighbor nodes connected to node k, and a_{lk} are the combination coefficients of neighbor nodes at node k such that

$$a_{lk} = 0 \text{ if } l \notin \mathcal{N}_k, \; a_{lk} > 0 \text{ if } l \in \mathcal{N}_k, \text{ and} \sum_{l \in \mathcal{N}_k} a_{lk} = 1. \quad (2)$$

As shown in Fig. 1, as the measurement data at each node and the unknown system are analog and each agent processes local data $\{d_k(i), \mathbf{x}_k(i)\}$ digitally, we need two ADCs in each agent. One concern is that as the number of agents increases, the power consumption will grow considerably when using high-resolution ADCs for each agent. This motivates us to quantize signals using few bits. Therefore, the problem we are interested in solving is how to design energy-efficient distributed learning algorithms that can cost-effectively operate with coarsely quantized signals.

III. Proposed DQA-LMS Algorithm

Let $\mathbf{x}_{k,Q} = Q_b(\mathbf{x}_k)$ denote the b-bit quantized output of an ADC at node k, described by a set of $2^b + 1$ thresholds $\mathcal{T}_b = \{\tau_0, \tau_1, \ldots, \tau_{2^b}\}$, such that $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{2^b} = \infty$, and the set of 2^b labels $\mathcal{L}_b = \{l_0, l_1, \ldots, l_{2^b-1}\}$ where $l_p \in (\tau_p, \tau_{p+1}]$, for $p \in [0, 2^b - 1]$ [18]. Let us assume that $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_{x_k})$, where $\mathbf{R}_{x_k} \in \mathbb{C}^{M \times M}$ is the covariance matrix of $\mathbf{x}_k$. We now use Bussgang's theorem [23] to derive a model for the quantized vector $\mathbf{x}_{k,Q}$, which we will use later to derive our DQA-LMS algorithm. Employing Bussgang's theorem, $\mathbf{x}_{k,Q}$ can be decomposed as

$$\mathbf{x}_{k,Q} = \mathbf{G}_{k,b}\mathbf{x}_k + \mathbf{q}_k, \quad (3)$$

where the quantization distortion $\mathbf{q}_k$ is uncorrelated with $\mathbf{x}_k$, and $\mathbf{G}_{k,b} \in \mathbb{R}^{M \times M}$ is a diagonal matrix described by

$$\mathbf{G}_{k,b} = \text{diag}(\mathbf{R}_{x_k})^{-\frac{1}{2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left[\exp(-\tau_j^2 \text{diag}(\mathbf{R}_{x_k})^{-1}) - \exp(-\tau_{j+1}^2 \text{diag}(\mathbf{R}_{x_k})^{-1})\right]. \quad (4)$$

Note that this signal decomposition is also applied to the desired signal, $d_{k,Q}$, which is the output of the second ADC in the system, and for the particular case that $\mathbf{R}_{x_k} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H] = \sigma_{x,k}^2 \mathbf{I}_M$, $\mathbf{G}_{k,b}$ becomes $g_{k,b}\mathbf{I}_M$. However, to minimize the mean square error (MSE) between $\mathbf{x}_k$ and $\mathbf{x}_{k,Q}$, we need to characterize the probability density function (PDF) of $\mathbf{x}_k$ to find the optimal quantization labels. Since the choice of labels based on the PDF is not practical, we assume the regressor $\mathbf{x}_k(i)$ is Gaussian, adapt the approach in [18] and approximate the thresholds and labels as follows:

1) We generate an auxiliary Gaussian random variable with unit variance and then use the Lloyd-Max algorithm [24], [25] to find a set of thresholds $\widetilde{\mathcal{T}}_b = \{\tau_1, \ldots, \tau_{2^b-1}\}$ and labels $\widetilde{\mathcal{L}}_b = \{\widetilde{l}_0, \ldots, \widetilde{l}_{2^b-1}\}$ that minimize the MSE between the unquantized and the quantized signals.
2) We complete the set of thresholds $\mathcal{T}_b$ by adding $\tau_0 = -\infty$ and $\tau_{2^b} = \infty$ to the set $\widetilde{\mathcal{T}}_b$.
3) We rescale the labels such that the variance of the auxiliary random variable is 1. To do this, we multiply each label in the set $\widetilde{\mathcal{L}}_b$ by

$$\alpha = \left(2 \sum_{j=0}^{2^b-1} \widetilde{l}_j^2 (\Phi(\sqrt{2\tau_{j+1}^2}) - \Phi(\sqrt{2\tau_j^2})\right)^{-1/2} \quad (5)$$

to produce a set of suboptimal labels $\mathcal{L}_b = \alpha \widetilde{\mathcal{L}}_b$, where $\Phi(.)$ is the cumulative distribution function (CDF) of a standard Gaussian random variable.

We generate these thresholds and labels offline to build $\mathbf{G}_{k,b}$ for the proposed DQA-LMS algorithm in what follows.

A. Derivation of DQA-LMS

We consider $\mathbf{x}_k(t)$ and $d_k(t)$ as the analog input and output of the unknown system $\mathbf{w}_o$ at node k. Let $\mathbf{x}_k(i)$ and $d_k(i)$ denote the high-precision sampled versions of $\mathbf{x}_k(t)$ and $d_k(t)$, and $\mathbf{x}_{k,Q}(i)$ and $d_{k,Q}(i)$ denote the coarsely quantized versions of $\mathbf{x}_k(i)$ and $d_k(i)$, respectively. We assume that the input signal at each node is Gaussian with zero mean and covariance matrix $\mathbf{R}_{x_k} = E[\mathbf{x}_k\mathbf{x}_k^H] = \sigma_{x,k}^2\mathbf{I}_M$ for $k = 1, 2, \ldots, N$. Using (3), we can decompose $\mathbf{x}_{k,Q}(i)$ and $d_{k,Q}(i)$ as

$$\mathbf{x}_{k,Q}(i) = g_{k,b}(i)\mathbf{x}_k(i) + \mathbf{q}_{x,k}(i), \quad (6)$$

$$\begin{aligned} d_{k,Q}(i) &= Q(d_k(i)) \approx g_{k,b}(i)d_k(i) + q_k(i) \\ &= g_{k,b}(i)\mathbf{w}_o^H\mathbf{x}_k(i) + \hat{q}_k(i), \end{aligned} \quad (7)$$

where $\hat{q}_k(i) = g_{k,b}(i)v_k(i) + q_k(i)$ and $g_{k,b}(i)$ are built from an estimate of $\mathbf{R}_{x_k}$ given by $\widehat{\mathbf{R}}_{x_k} = \mathbf{x}_k\mathbf{x}_k^H$ [26] that depends on the choice of $\mathbf{x}_k$ due to (1). Because the adaptive algorithm receives a quantized signal, $\mathbf{x}_{k,Q}$, and the signal is assumed to be wide-sense stationary, at each time instant, we estimate $\sigma_{x,k}^2$ using the variance of the received input, $\sigma_{x_{k,Q}}^2$ and the distortion factor of the b-bit quantization, $\rho_{k,b}$, such that $\sigma_{x,k}^2 \approx \sigma_{x_{k,Q}}^2 + \rho_{k,b}$, where $\rho_{k,b} \approx \frac{\pi\sqrt{3}}{2} 2^{-2b}$ [19] for a Gaussian signal using non-uniform quantization to obtain the scalar $g_{k,b}(i)$.

We show next that a learning algorithm based directly on (7) is biased for estimating $\mathbf{w}_o$, and show how to correct for this bias. For this, let $\beta_k(i)$ be a coefficient to be chosen shortly, and define $\hat{d}_k(i) = \beta_k(i)\mathbf{w}_k^H(i-1)\mathbf{x}_{k,Q}(i)$ and construct an MSE cost function as described by

$$\begin{aligned} J_k(\mathbf{w}_k(i)) &= \mathbb{E}[|e_{k,Q}(i)|^2] = \mathbb{E}[|d_{k,Q}(i) - \hat{d}_k(i)|^2] \\ &= \mathbb{E}[|d_{k,Q}(i) - \beta_k(i)\mathbf{w}_k^H(i-1)\mathbf{x}_{k,Q}(i)|^2], \end{aligned} \quad (8)$$

which depends only on the observed quantized quantities $d_{k,Q}(i)$ and $\mathbf{x}_{k,Q}(i)$. For $\beta_k(i) = 1$ as in DLMS, the quantization of $d_k(i)$ would result in biased estimates of $\mathbf{w}_o$. In the following we show how to optimally choose $\beta_k(i)$ to reduce the bias. The proposed gradient-descent recursion to perform distributed learning based on (8) is described by

$$\mathbf{h}_k(i) = \mathbf{w}_k(i-1) - \mu_k \nabla J_k(\mathbf{w}_k(i-1)). \quad (9)$$

To compute the gradient of (9), we write the error in (8) as

$$\begin{aligned} e_{k,Q}(i) &= d_{k,Q}(i) - \beta_k(i)\mathbf{w}_k^H(i-1)\mathbf{x}_{k,Q}(i) \\ &= g_{k,b}(i)\mathbf{w}_o^H\mathbf{x}_k(i) + \hat{q}_k(i) - \beta_k(i)\mathbf{w}_k^H(i-1) \\ &\quad (g_{k,b}(i)\mathbf{x}_k(i) + \mathbf{q}_{x,k}(i)) \\ &= g_{k,b}(i)(\mathbf{w}_o^H - \beta_k(i)\mathbf{w}_k^H(i-1))\mathbf{x}_k(i) \\ &\quad - \beta_k(i)\mathbf{w}_k^H(i-1)\mathbf{q}_{x,k}(i) + \hat{q}_k(i). \end{aligned} \quad (10)$$

We assume that $\mathbf{R}_{x_k} = \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H] = \sigma_{x,k}^2\mathbf{I}_M$ and $\mathbf{R}_{q,k} = \mathbb{E}[\mathbf{q}_{x,k}\mathbf{q}_{x,k}^H] = \sigma_{q,k}^2\mathbf{I}_M$. Substituting (10) in (8) and taking

TABLE I
COMPUTATIONAL COMPLEXITY PER TIME INSTANT

Task	Multiplications	Additions	Divisions	Exponentiations
$g_{k,b}(i) = \frac{1}{\sqrt{\sigma_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}}[\exp(\frac{-\tau_j^2}{\sigma_{x_k}^2}) - \exp(\frac{-\tau_{j+1}^2}{\sigma_{x_k}^2})]$	$2^{b+1}+1$	2^b-1	2^b+1	2^b
$\beta_k(i) = \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2+\sigma_{q,k}^2}$	2	1	1	0
$\hat{d}_{k,Q}(i) = \beta_k(i)\mathbf{w}_k^H(i)\mathbf{x}_{k,Q}(i)$	$M+1$	$M-1$	0	0
$e_{k,Q}(i) = d_{k,Q}(i) - \hat{d}_{k,Q}(i)$	0	1	0	0
$\mathbf{h}_k(i+1) = \mathbf{w}_k(i) + \mu_k\beta_k(i)e_{k,Q}^*(i)\mathbf{x}_{k,Q}(i)$	$M+2$	M	0	0
$\mathbf{w}_k(i+1) = \sum_{l\in\mathcal{N}_k} a_{lk}\mathbf{h}_l(i+1)$	n_kM	n_kM	0	0
Total (at node k)	$(2+n_k)M+2^{b+1}+6$	$(2+n_k)M+2^b$	2^b+2	2^b

the expected value of (9), we have

$$\mathbb{E}[\mathbf{h}_k(i)] = [\mathbf{I}_M - \mu_k g_{k,b}^2(i)\beta_k(i)\mathbf{R}_{x_k} - \mu_k g_{k,b}(i)\beta_k(i)\mathbf{R}_{q,k}] \cdot \mathbb{E}[\mathbf{w}_k(i-1)] + \mu g_{k,b}^2(i)\mathbf{R}_{x_k}\mathbf{w}_o^H. \quad (11)$$

Substituting the values of $\mathbf{R}_{x_k}$ and $\mathbf{R}_{q,k}$ and taking the limit on (11), we obtain

$$\lim_{i\to+\infty} \mathbb{E}[\mathbf{h}_k(i)] = \frac{1}{\beta_k(i)} \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2}\mathbf{w}_o. \quad (12)$$

We conclude that the solution is unbiased if we choose

$$\beta_k(i) = \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2}. \quad (13)$$

The gradient of $|e_{k,Q}(i)|^2$ with respect to $\mathbf{w}_k^H$ is $\nabla J_k(\mathbf{w}_k(i-1)) = -\frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2+\sigma_{q,k}^2}\mathbf{x}_{k,Q}(i)e_{k,Q}^*(i)$. After organizing the terms of the gradient, we obtain the DQA-LMS algorithm:

$$\mathbf{h}_k(i) = \mathbf{w}_k(i-1) + \mu_k \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2}\mathbf{x}_{k,Q}(i)e_{k,Q}^*(i),$$

$$\mathbf{w}_k(i) = \sum_{l\in\mathcal{N}_k} a_{lk}\mathbf{h}_l(i), \quad (14)$$

$$e_{k,Q}(i) = d_{k,Q}(i) - \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2}\mathbf{w}_k^H(i-1)\mathbf{x}_{k,Q}(i)\,,$$

$$g_{k,b}(i) = \frac{1}{\sqrt{\sigma_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left[\exp\left(\frac{-\tau_j^2}{\sigma_{x_k}^2}\right) - \exp\left(\frac{-\tau_{j+1}^2}{\sigma_{x_k}^2}\right)\right], \quad (15)$$

and $\sigma_{x,k}^2 \approx \sigma_{x_{k,Q}}^2 + \rho_{k,b}$. The scalar $g_{k,b}$ can be computed offline when $\mathbf{R}_{x,k}$ is known and wide-sense stationary and must be estimated online when $\mathbf{R}_{x,k}$ is unknown or non-stationary.

B. Computational Complexity and Energy Consumption

Table I shows the computational complexity of the DQA-LMS algorithm in terms of the number of multiplications and additions at node k per time instant, where n_k is the number of neighbor nodes connected to node k. At each time instant, DQA-LMS performs a few more operations ($\approx O(2^b)$) than DLMS. Note that we compute $g_{k,b}(i)$ online since this is more appropriate for non-stationary input data. However, one can compute $\mathbf{G}_{k,b}$ offline if an estimate of $\mathbf{R}_{x_k}$ in (4) is available.

However, the extra complexity of DQA-LMS allows the system to work in a more energy-efficient way. In order to assess the power savings by low resolution quantization, we consider a network with N nodes in which each node uses two ADCs. The power consumption of each ADC is $P_{ADC}(b) = cB2^b$ [27], where B is the bandwidth (related to the sampling rate), b is the number of quantization bits of the ADC, and c is the power consumption per conversion step. Therefore, the total power consumption of the ADCs in the network is

$$P_{ADC,T}(b) = 2NcB2^b \qquad \text{(watts)}. \quad (16)$$

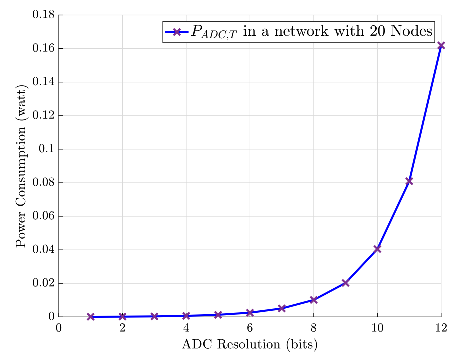


Fig. 2. Power consumption of the ADCs in an adaptive IoT network.

Fig. 2 shows an example of the total power consumption of ADCs in a narrowband IoT (NB-IoT) network running diffusion adaptation consisting of 20 nodes with bandwidth $B = 200\,\text{kHz}$ [28] and considering the power consumption per conversion step of each ADC, $c = 494$ fJ, as in [29].

IV. Analysis of DQA-LMS

In this section, we find sufficient conditions for all local estimates to converge in the mean to the unknown parameter vector $\mathbf{w}_o$ by using the evolution of the weight error vectors [4]. Let us consider the global quantities of the network: $\mathbf{W}_o \triangleq [\mathbf{w}_o, \ldots, \mathbf{w}_o]_{(NM\times 1)}$, $\mathbf{d}_Q(i) \triangleq [d_{1,Q}(i), \ldots, d_{N,Q}(i)]^T$, $\mathbf{v}_i \triangleq [v_1(i), \ldots, v_N(i)]^T$, $\mathbf{X}_Q(i) \triangleq \text{diag}[\mathbf{x}_{1,Q}^T(i), \ldots, \mathbf{x}_{N,Q}^T(i)]$.

Using these quantities, the global form of (1) is given by $\mathbf{d}_Q(i) = \mathbf{W}_o^H\mathbf{X}_Q(i) + \mathbf{v}(i)$. Defining $\mathbf{B}(i)$, $\mathbf{W}(i)$ and $\mathbf{H}(i)$ as the global quantities for, respectively, $\beta_k(i)$, $\mathbf{h}_k(i)$ and $\mathbf{w}_k(i)$, we can express (14) as

$$\mathbf{H}(i) = \mathbf{W}(i-1) + \mathbf{D}\mathbf{B}(i)\mathbf{X}_Q(i)(\mathbf{d}_Q(i) - \mathbf{B}(i)\mathbf{W}^H(i-1)\mathbf{X}_Q(i))^*, \quad \mathbf{W}(i) = \mathbf{C}\mathbf{H}(i), \quad (17)$$

which can be written in a compact form as

$$\mathbf{W}(i) = \mathbf{C}\mathbf{W}(i-1) + \mathbf{C}\mathbf{D}\mathbf{B}(i)\mathbf{X}_Q(i)(\mathbf{d}_Q(i) - \mathbf{B}(i)\mathbf{W}^H(i-1)\mathbf{X}_Q(i))^*, \quad (18)$$

where $\mathbf{D} \triangleq \text{diag}\{\mu_1 I_M, \ldots, \mu_N I_M\}$ and $\mathbf{C}$ is an $MN \times MN$ matrix based on the combination coefficients, a_{lk}, defined as

$$\mathbf{A} \triangleq [a_{ij}]\,, \qquad \mathbf{C} \triangleq \mathbf{A} \otimes \mathbf{I}_M. \quad (19)$$

Using the independence assumption [4] that states that $\mathbf{x}_{k,Q}(i)$ and $v_k(i)$ are i.i.d. in time and space with $\sigma^2_{v,k} = \mathbb{E}[|v_k(i)|^2]$, and $v_k(i)$ is independent of $\mathbf{x}_{k,Q}(i)$, we define the weight error vector, $\widetilde{\mathbf{w}}_k(i)$ and its global vector $\widetilde{\mathbf{w}}(i)$ as

$$\widetilde{\mathbf{W}}(i) \triangleq \mathbf{W}_o - \mathbf{W}(i). \tag{20}$$

Note that using diffusion combination policies for a_{lk}, we have $\mathbf{CW}_o = \mathbf{W}_o$ [4]. Subtracting $\mathbf{W}_o$ from the left-hand side and $\mathbf{CW}_o$ from the right-hand side of (18), we have

$$\begin{aligned}\widetilde{\mathbf{W}}(i) &= \mathbf{CW}_o - \mathbf{CW}(i-1) - \mathbf{CDB}(i)\mathbf{X}_Q(i)\\ &= \mathbf{C}\widetilde{\mathbf{W}}(i-1) - \mathbf{CDB}(i)\mathbf{X}_Q(i)\\ &\quad (\mathbf{X}^*_Q(i)\mathbf{B}(i)^*\widetilde{\mathbf{W}}(i-1) + \mathbf{v}^*(i))\\ &= \mathbf{C}(\mathbf{I}_{MN} - \mathbf{DB}(i)\mathbf{X}_Q(i)\mathbf{X}^*_Q(i)\mathbf{B}(i)^*)\widetilde{\mathbf{W}}(i-1)\\ &\quad - \mathbf{CDB}(i)\mathbf{X}_Q(i)\mathbf{v}^*(i).\end{aligned} \tag{21}$$

Taking the expectation of both sides of (21), we have

$$\mathbb{E}[\widetilde{\mathbf{W}}(i)] = C(\mathbf{I}_{MN} - \mathbf{DR}_Q)\mathbb{E}[\widetilde{\mathbf{W}}(i-1)], \tag{22}$$

where $C \triangleq \mathbb{E}[\mathbf{C}]$, $\mathbf{R}_Q \triangleq \mathrm{diag}\{\mathbf{R}_{1,Q}, \ldots, \mathbf{R}_{N,Q}\}$ and $\mathbf{R}_{k,Q} = \mathbb{E}[\mathbf{B}_k(i)\mathbf{x}_{k,Q}(i)\mathbf{x}^*_{k,Q}(i)\mathbf{B}_k(i)^*]$.

To ensure stability of the recursion in (22) with the independence assumption and using combinations that satisfy (2), there exist sufficiently small step-sizes $\mu_k < \mu_{max}$ such that

$$\|\mathbb{E}[\widetilde{\mathbf{W}}(i)]\|_{bl_\infty} \leq \|C\|_{bl_\infty}.\|\mathbf{E}_i\|_{bl_\infty}.\|\mathbb{E}[\widetilde{\mathbf{w}}(i-1)]\|_{bl_\infty}, \tag{23}$$

where $\|.\|_{bl_\infty}$ denotes the block maximum norm [30] and $\mathbf{E}_i = \mathbf{I}_{MN} - \mathbf{DR}_Q$. In order for DQA-LMS to converge, we hold (23) such that $\|\mathbf{E}\|_{bl_\infty} < 1$ and $\|C_i\|_{bl_\infty} \leq 1$ for all $i \geq 0$. It is proven in [30] that possibly random, time-varying convex combinations generated by ATC or CTA diffusion algorithms ensure $\|C_i\|_{bl_\infty} \leq 1$. Therefore, to find sufficient conditions on step-sizes, we must have $\mathbf{I}_{MN} - \mathbf{DR}_Q < 1$.

We now employ the eigenvalue decomposition $\mathbf{R}_{k,Q} = \mathbf{\Phi}_{k,Q}\mathbf{\Lambda}_{k,Q}\mathbf{\Phi}^H_{k,Q}$, where $\mathbf{\Lambda}_{k,Q}$ is an $M \times M$ diagonal matrix consisting of the eigenvalues $\{\lambda_{(k,Q)_1}, \ldots, \lambda_{(k,Q)_M}\}$ of $\mathbf{R}_{k,Q}$, and the matrix $\mathbf{\Phi}_{k,Q}$ is an $M \times M$ square matrix whose columns are the eigenvectors $\{\phi_{(k,Q)_1}, \ldots, \phi_{(k,Q)_M}\}$ of $\mathbf{R}_{k,Q}$ associated with these eigenvalues. We define $\mathbf{\Phi}_Q \triangleq \mathrm{diag}\{\mathbf{\Phi}_{1,Q}, \ldots, \mathbf{\Phi}_{N,Q}\}$ and $\mathbf{\Lambda}_Q \triangleq \mathrm{diag}\{\mathbf{\Lambda}_{1,Q}, \ldots, \mathbf{\Lambda}_{N,Q}\}$. Since $\mathbf{\Phi}^H_Q\mathbf{D}\mathbf{\Phi}_Q = \mathbf{D}$, the condition on the step size can be written as $\|\mathbf{I}_{MN} - \mathbf{D\Lambda}_Q\|_\infty < 1$, which yields

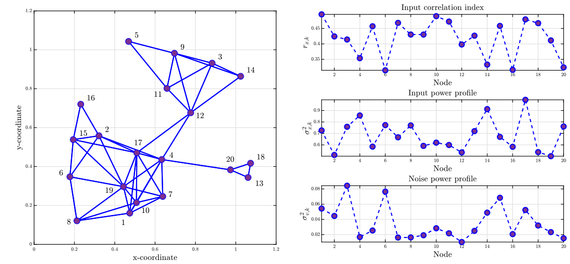
$$\begin{aligned}\|\mathbf{I}_{MN} - \mathbf{D\Lambda}_Q\|_\infty &= \max_{1\leq k\leq N} \|\mathbf{I}_M - \mu_k\mathbf{\Lambda}_{k,Q}\|\\ &= \max_{1\leq k\leq N}\max_{1\leq m\leq M} |1 - \mu_k\lambda_{(k,Q)_m}| < 1,\end{aligned}$$

where $\lambda_{(k,Q)_m}$ is the mth diagonal eigenvalue of $\mathbf{R}_{k,Q}$. Therefore, the stability condition for DQA-LMS is given by

$$0 < \mu_k < \frac{2}{\lambda_{\max}(\mathbf{R}_{k,Q})} \quad \text{for all } k = 1, 2, \ldots, N. \tag{24}$$

V. Simulation Results

In this section, we assess the performance of the DQA-LMS algorithm for a parameter estimation problem in an IoT network with $N = 20$ nodes. The impulse response of the unknown system has $M = 8$ taps, is generated randomly and normalized to one. The input signals $\mathbf{x}_k(i)$ at each node are generated by passing a white Gaussian noise process with variance $\sigma^2_{x,k}$ through a first order autoregressive model with transfer function $\frac{1}{1-r_{x,k}z^{-1}}$ where $r_{x,k} \in (0.3, 0.5)$ are the correlation coefficients and quantized using Lloyd-Max quantization scheme to generate $\mathbf{x}_{k,Q}(i)$. The noise samples of each node are drawn from a zero mean white Gaussian process with variance $\sigma^2_{v,k}$. Fig. 3 plots the network details.



(a) Distributed network structure (b) Variances and correlation coefficients

Fig. 3. A wireless network with $N = 20$ nodes.

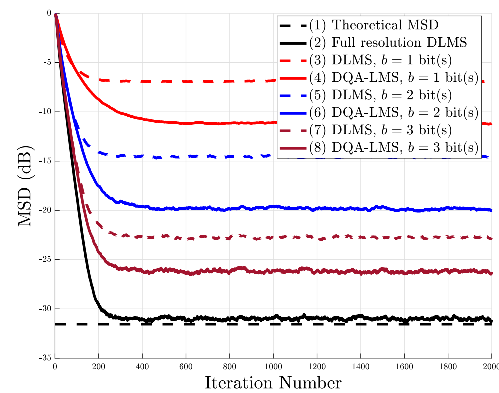


Fig. 4. The MSD curves for the DLMS and DQA-LMS algorithms.

The simulated mean-square deviation (MSD) learning curves are obtained by ensemble averaging over 100 independent trials. We choose the same step sizes for all agents, i.e., $\mu_k = 0.05$. The combining coefficients a_{lk} are computed by the Metropolis rule. The evolution of the ensemble-average learning curves, $\frac{1}{N}\mathbb{E}[\|\widetilde{\mathbf{w}}_i\|^2]$, for the ATC diffusion strategy using different numbers of bits is assessed. The theoretical MSD of the DLMS with the same step size μ and the Metropolis rule applied to a_{lk} is approximated by $\frac{\mu M}{N^2}\sum_{k=1}^{N}\sigma^2_{v,k}$ [5] and shown by curve 1. Curve 2 shows the standard DLMS performance assuming full resolution ADCs to perform system identification. Curves 3, 5 and 7 show the MSD evolution of the standard DLMS with low resolution signals coarsely quantized with 1, 2 and 3 bits, respectively. Curves 4, 6 and 8 show the MSD performance of the proposed DQA-LMS algorithm that improves the error measurement confronted with coarsely quantized signals. The performance of the proposed DQA-LMS algorithm is closer to the DLMS while it reduces about 90% of the power consumption by ADCs in the network (see Fig. 2).

VI. Conclusion

In this letter, we have proposed an energy-efficient framework for distributed learning and developed the DQA-LMS algorithm using low-resolution ADCs for adaptive IoT networks. DQA-LMS has comparable computational cost to the full-resolution DLMS algorithm while it enormously reduces the power consumption of the ADCs in the network. Simulations have shown the close performance of DQA-LMS to the DLMS algorithm despite dealing with coarsely quantized signals.

REFERENCES

[1] J. B. Predd, S. B. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 56–69, Jul. 2006.

[2] M. M. Rana, W. Xiang, and E. Wang, "IOT-based state estimation for microgrids," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1345–1346, Apr. 2018.

[3] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[4] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[5] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 155–171, May 2013.

[6] S. Xu, R. C. de Lamare, and H. V. Poor, "Distributed estimation over sensor networks based on distributed conjugate gradient strategies," *IET Signal Process.*, vol. 10, no. 3, pp. 291–301, 2016.

[7] T. G. Miller, S. Xu, R. C. de Lamare, and H. V. Poor, "Distributed spectrum estimation based on alternating mixed discrete-continuous adaptation," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 551–555, Apr. 2016.

[8] S. Xu, R. C. de Lamare, and H. V. Poor, "Distributed compressed estimation based on compressive sensing," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1311–1315, Sep. 2015.

[9] T. G. Miller, S. Xu, R. C. de Lamare, V. H. Nascimento, and Y. Zakharov, "Sparsity-aware distributed conjugate gradient algorithms for parameter estimation over sensor networks," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, 2015, pp. 1556–1560.

[10] S. Xu, R. C. de Lamare, and H. V. Poor, "Adaptive link selection algorithms for distributed estimation," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, pp. 1–22, 2015.

[11] S. Zhang and W. X. Zheng, "Distributed separated-decorrelation LMS algorithms over sensor networks with noisy inputs," *IEEE Trans. Signal Process.*, vol. 68, pp. 4163–4177, 2020.

[12] Y. Yu, H. Zhao, R. C. de Lamare, Y. Zakharov, and L. Lu, "Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1566–1581, Mar. 2019.

[13] C. Han, J. M. Jornet, E. Fadel, and I. F. Akyildiz, "A cross-layer communication module for the Internet of Things," *Comput. Netw.*, vol. 57, no. 3, pp. 622–633, 2013.

[14] I. Utlu, O. F. Kilic, and S. S. Kozat, "Resource-aware event triggered distributed estimation over adaptive networks," *Digit. Signal Process.*, vol. 68, pp. 127–137, 2017.

[15] A. Mezghani and J. A. Nossek, "Power efficiency in communication systems from a circuit perspective," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2011, pp. 1896–1899.

[16] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[17] L. T. N. Landau and R. C. de Lamare, "Branch-and-bound precoding for multiuser MIMO systems with 1-bit quantization," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 770–773, Dec. 2017.

[18] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution adcs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[19] A. Mezghani, M.-S. Khoufi, and J. A. Nossek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. ITG/IEEE WSA*, Vienna, Austria, 2007, pp. 1–5.

[20] L. T. N. Landau, M. Dörpinghaus, R. C. de Lamare, and G. P. Fettweis, "Achievable rate with 1-bit quantization and oversampling using continuous phase modulation-based sequences," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 7080–7095, Oct. 2018.

[21] Z. Shao, R. C. de Lamare, and L. T. N. Landau, "Iterative detection and decoding for large-scale multiple-antenna systems with 1-bit ADCS," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 476–479, Jun. 2018.

[22] Z. Shao, L. Landau, and R. C. de Lamare, "Adaptive RLS channel estimation and SIC for large-scale antenna systems with 1-bit ADCs," in *Proc. 22nd Int. ITG Workshop Smart Antennas.*, 2018, pp. 1–4.

[23] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Research Lab. Electron MIT, Cambridge, MA, USA, Tech. Rep. 216, 1952.

[24] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[25] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[26] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[27] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. Inf. Theory Appl. Workshop.*, 2015, pp. 191–198.

[28] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "NB-IOT system for M2M communication," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2016, pp. 1–5.

[29] H. Chung, A. Rylyakov, Z. T. Deniz, J. Bulzacchelli, G. Wei, and D. Friedman, "A 7.5-GS/s 3.8-ENOB 52-mW flash ADC with clock duty cycle control in 65 nm CMOS," in *Proc. Symp. VLSI Circuits*, 2009, pp. 268–269.

[30] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810, Sep. 2010.