ENERGY-EFFICIENT DISTRIBUTED LEARNING WITH ADAPTIVE BIAS COMPENSATION FOR COARSELY QUANTIZED SIGNALS

Alireza Danaee[⋆], *Rodrigo C. de Lamare*[⋆,†] *and* *Vítor H. Nascimento*[‡]

[⋆] Centre for Telecommunications Studies, Pontifical Catholic University of Rio de Janeiro, Brazil
[†] Department of Electronic Engineering, University of York, United Kingdom
[‡] Department of Electronic Systems Engineering, University of São Paulo, Brazil
{alireza;delamare}@cetuc.puc-rio.br, vitor@lps.usp.br

ABSTRACT

In this work, we consider an energy-efficient distributed learning framework using low-resolution ADCs and coarsely quantized signals for Internet of Things (IoT) networks. We develop an adaptive bias compensation strategy to improve the distributed quantization-aware least-mean square (DQA-LMS) algorithm and propose the Adaptive DQA-LMS (AdDQA-LMS) algorithm that can learn parameters in a distributed network with an energy-efficient fashion using signals quantized with few bits while requiring a low computational cost. Simulations assess the AdDQA-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 AdDQA-LMS algorithm.

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

1. 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], and robust techniques against interference and noise [11]. 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 quite limited. A distributed quantization-aware algorithm was proposed in [12] to reduce the power consumption of analog-to-digital converters (ADCs) in the adaptive IoT networks in an energy-efficient framework.

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] and from a circuit perspective on analog-to-digital converters (ADCs) and decoders [14]. Reducing the number of bits used to represent digital samples can greatly decrease the energy consumption by ADCs [15]. 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 [16–20].

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 an adaptive distributed quantization-aware least-mean square (AdDQA-LMS) algorithm that employs an adaptive bias compensation strategy (as opposed to the fixed compensation of [12]) and that can learn parameters in an energy-efficient way using signals quantized using few bits with a low computational cost. Simulations assess the AdDQA-LMS algorithm against existing techniques for a distributed parameter estimation task with IoT devices.

This paper is structured as follows: Section 2 introduces the signal model and states the problem, whereas Section 3 details the proposed AdDQA-LMS algorithm. Section 4 shows and discusses the simulation results and Section 5 draws the conclusions of this work.

2. SIGNAL MODEL AND PROBLEM STATEMENT

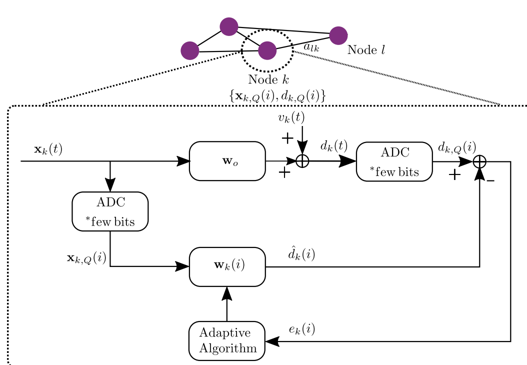


Fig. 1. A distributed adaptive IoT network

We consider an IoT network with 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

This work was supported in part by CNPq, CAPES, FAPERJ and by the ELIOT Project (FAPESP 2018/12579-7, ANR-18-CE40-0030).

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

where $\mathbf{w}_o \in \mathbb{C}^{M\times 1}$ is the parameter vector that the agents must estimate, $\mathbf{x}_k(i) = [x_k(i), x_k(i-1), \ldots, x_k(i-M+1)]^T \in \mathbb{C}^{M\times 1}$ is the regressor and $v_k(i)$ is Gaussian noise with zero mean and variance $\sigma_{v,k}^2$ 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 their neighborhood, the parameter vector with local estimates $\mathbf{w}_k(i)$ is updated. The ATC distributed LMS (DLMS) algorithm uses 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

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

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.

3. PROPOSED ADDQA-LMS ALGORITHM

3.1. Signal Decomposition

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 $\mathcal{T}_b = \{\tau_0, \tau_1, ..., \tau_{2^b}\}$, such that $-\infty = \tau_0 < \tau_1 < ... < \tau_{2^b} = \infty$, and the set of 2^b labels $\mathcal{L}_b = \{l_0, l_1, ..., l_{2^b-1}\}$ where $l_p \in (\tau_p, \tau_{p+1}]$, for $p \in [0, 2^b-1]$ [17]. 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 [21] to derive a model for the quantized vector $\mathbf{x}_{k,Q}$, which we will use later to derive our AdDQA-LMS algorithm. Employing Bussgang's theorem, $\mathbf{x}_{k,Q}$ can be decomposed as

$$\mathbf{x}_{k,Q} = \mathbf{G}_{k,b}\mathbf{x}_k + \mathbf{q}_k, \qquad (2)$$

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

$$\mathbf{G}_{k,b} = \text{diag}(\mathbf{R}_{x_k})^{-\frac{1}{2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left[\exp(-\tau_j^2 \text{diag}(\mathbf{R}_{x_k})^{-1}) - \exp(-\tau_{j+1}^2 \text{diag}(\mathbf{R}_{x_k})^{-1})\right], \qquad (3)$$

where $\text{diag}(\mathbf{R}_{x_k})$ is the $N \times N$ diagonal matrix whose entries are the N diagonal elements of the matrix $\mathbf{R}_{x_k}$. Note that, as a simplifying approximation, we also apply this signal decomposition 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 $g_{k,b}\mathbf{I}_M$ where

$$g_{k,b} = \frac{1}{\sqrt{\sigma_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left(e^{-\frac{\tau_j^2}{\sigma_{x_k}^2}} - e^{-\frac{\tau_{j+1}^2}{\sigma_{x_k}^2}}\right). \qquad (4)$$

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. Because choosing these labels based on such PDF is ineffective in practice (since the PDFs are difficult to estimate), we assume the regressor $\mathbf{x}_k(i)$ is Gaussian, then adapt the approach in [17] and approximate the thresholds and labels offline later on with the AdDQA-LMS algorithm.

3.2. The DQA-LMS Algorithm

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 $k = 1, 2, ..., N$. Using (2), we can decompose $\mathbf{x}_{k,Q}(i)$ and $d_{k,Q}(i)$ as

$$\mathbf{x}_{k,Q}(i) = g_{k,b}(i)\mathbf{x}_k(i) + \mathbf{q}_{x,k}(i), \qquad (5)$$

$$\begin{aligned} d_{k,Q}(i) &= Q(d_k(i)) \approx g_{k,b}(i)d_k(i) + q_k(i) \\ &= g_{k,b}(i)\mathbf{w}_o^H\mathbf{x}_k(i) + \hat{q}_k(i), \end{aligned} \qquad (6)$$

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 $\widehat{\mathbf{R}}_{x_k} = \mathbf{x}_k\mathbf{x}_k^H$ that depends on the choice of $\mathbf{x}_k$ due to (1).

It has been shown in [12] that a learning algorithm based on (6) is biased for estimating $\mathbf{w}_o$, and an approach to correct for this bias was devised. To this end, let $\beta_k(i)$ denote the bias compensation coefficient, 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

$$\begin{aligned} J_k(\mathbf{w}_k(i)) &= \mathbb{E}[|e_{k,Q}(i)|^2] = \mathbb{E}[|d_{k,Q}(i) - \hat{d}_k(i)|^2] \\ &= \mathbb{E}[|d_{k,Q}(i) - \beta_k(i)\mathbf{w}_k^H(i-1)\mathbf{x}_{k,Q}(i)|^2], \end{aligned} \qquad (7)$$

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 $\mathbf{w}_o$.

The DQA-LMS solution to $J_k(\mathbf{w}_k(i))$ is given by [12]

$$\begin{aligned} \mathbf{h}_k(i) &= \mathbf{w}_k(i-1) + \mu_k\beta_k(i)\mathbf{x}_{k,Q}(i)e_{k,Q}^*(i), \\ \mathbf{w}_k(i) &= \sum_{l\in\mathcal{N}_k} a_{lk}\mathbf{h}_l(i). \end{aligned} \qquad (8)$$

This solution is unbiased if we choose [12]

$$\beta_k(i) = \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2}, \qquad (9)$$

where $\sigma_{q,k}^2$ is the variance of the quantization noise. Note that $g_{k,b}$ can be computed offline when $\mathbf{R}_{x_k}$ is wide-sense stationary and known or must be estimated online when $\mathbf{R}_{x_k}$ is unknown and non-stationary.

3.3. Proposed Adaptive Bias Compensation

In this section, we propose an adaptive strategy to compute the bias compensation term $\beta_k(i)$, which uses an approximation for $\mathbf{R}_{x_k}$ by the instantaneous values of the input vector. A simple approach to estimate $\mathbf{R}_{x_k}$ is obtained by employing the instantaneous values of $\mathbf{x}_k\mathbf{x}_k^H$ as follows

$$\widehat{\mathbf{R}}_{x_k}(i) = \mathbf{x}_k(i)\mathbf{x}_k^H(i) \qquad M \times M \tag{10}$$

and consequently for estimating $\mathbf{R}_{x_k,Q}(i)$

$$\widehat{\mathbf{R}}_{x_k,Q}(i) = \mathbf{x}_{k,Q}(i)\mathbf{x}_{k,Q}^H(i) \qquad M \times M. \tag{11}$$

Since we do not have access to the instantaneous values of $\mathbf{x}_k(i)$ and process $\mathbf{x}_{k,Q}(i)$, we use the following estimate

$$\widehat{\mathbf{R}}_{x_k}(i) = \widehat{\mathbf{R}}_{x_k,Q}(i) + \widehat{\sigma}_{q,k}^2\mathbf{I}_M. \tag{12}$$

Note that we estimate $\mathbf{R}_{x_k}$ online to form the diagonal entries of the matrix $\mathbf{G}_{k,b}(i)$ considering (3) and since $\mathbf{G}_{k,b}(i)$ is a diagonal matrix whose entries are built based on the diagonal entries of the covariance matrix $\mathbf{R}_{x_k}$, we only consider the diagonal entries of $\widehat{\mathbf{R}}_{x_k}(i)$ which can be shown as follows:

$$\widehat{\mathbf{R}}_{x_k}(i) = \begin{bmatrix} \widehat{r}_{x_k}(i) & & \\ & \ddots & \\ & & \widehat{r}_{x_k}(i-M+1) \end{bmatrix},$$

where $\widehat{r}_{x_k}(i) = x_{k,Q}(i)x_{k,Q}^*(i) + \widehat{\sigma}_{q,k}^2$.

We generate a set of thresholds $\mathcal{T}_b$ and labels $\mathcal{L}_b$, and estimate the variance of the quantization noise, $\sigma_{q,k}^2$ (as it is not accessible in practice) with the following offline procedure.

1. We generate an auxiliary Gaussian random variable, $\mathbf{x}_{aux}$, with unit variance and then use the Lloyd-Max algorithm [22], [23] to find a set of thresholds $\widetilde{\mathcal{T}}_b = \{\tau_1, \ldots, \tau_{2^b-1}\}$ and labels $\mathcal{L}_b = \{l_0, \ldots, l_{2^b-1}\}$ that minimize the MSE between the unquantized and the quantized signals.
2. We quantize $\mathbf{x}_{aux}$ using $\widetilde{\mathcal{T}}_b$ and $\mathcal{L}_b$, generate the quantized signal $\mathbf{x}_{aux,Q}$, and estimate the variance of the quantization noise, $\sigma_{q,k}^2$ with the subtraction of the variance of the quantized auxiliary signal from the variance of the auxiliary signal
$$\widehat{\sigma}_{q,k}^2 = \sigma_{\mathbf{x}_{aux}}^2 - \sigma_{\mathbf{x}_{aux,Q}}^2. \tag{13}$$
3. We wrap up the set of thresholds $\mathcal{T}_b$ by adding $\tau_0 = -\infty$ and $\tau_{2^b} = \infty$ to $\widetilde{\mathcal{T}}_b$.

Table 1 summarizes the AdDQA-LMS algrithm.

3.4. Energy Consumption

As $\boldsymbol{\beta}_k(i)$ in AdDQA-LMS is an $M \times M$ diagonal matrix instead of a single scalar $\beta_k(i)$ in DQA-LMS, we need more operations to generate the error and the intermediate estimate, however, the computational complexity still remains in the order of $O(M)$. The extra complexity of the DQA-LMS which is detailed in [12] allows the system to work in a more energy-efficient way and the adaptive strategy of the AdDQA-LMS enables the algorithm to be robust against variations and imprecise knowledge of $\mathbf{R}_{x_k}$. 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 $P_{ADC}(b) = cB2^b$ [24], 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

$$P_{ADC,T}(b) = 2NcB2^b \qquad \text{(watts)}. \tag{14}$$

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 $B = 200\,\text{kHz}$ [25] and considering the power consumption per conversion step of each ADC, $c = 494$ fJ, as in [26].

Table 1. Pseudo code of AdDQA-LMS algorithm

Initialization:

$\mathbf{w}_k(-1) = 0$ for each node k

Generate $\mathcal{T}_b$ and $\mathcal{L}_b$

Compute $\widehat{\sigma}_{q,k}^2$ from (13)

Define $\boldsymbol{\beta}_k(-1) = \text{diag}(b_k(-1), \ldots, b_k(-M)) = \mathbf{0}$

At each time instant i and node k

Receive data

$d_{k,Q}(i)$

$\mathbf{x}_{k,Q}(i) = [x_{k,Q}(i), x_{k,Q}(i-1), \ldots, x_{k,Q}(i-M+1)]^T$

Repeat

$\widehat{r}_{x_k}(i) = x_{k,Q}(i)x_{k,Q}^*(i) + \widehat{\sigma}_{q,k}^2$

$g_{k,b}(i) = \frac{1}{\sqrt{\widehat{r}_{x_k}(i)}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} \left(e^{-\frac{\tau_j^2}{\widehat{r}_{x_k}(i)}} - e^{-\frac{\tau_{j+1}^2}{\widehat{r}_{x_k}(i)}} \right)$

$b_k(i) = \frac{g_{k,b}(i)\widehat{r}_{x_k}(i)}{g_{k,b}(i)\widehat{r}_{x_k}(i) + \widehat{\sigma}_{q,k}^2}$, **update** $\boldsymbol{\beta}_k(i)$

$e_{k,Q}(i) = d_{k,Q}(i) - \mathbf{w}_k^H(i-1)\boldsymbol{\beta}_k(i)\mathbf{x}_{k,Q}(i)$

$\mathbf{h}_k(i) = \mathbf{w}_k(i-1) + \mu_k\boldsymbol{\beta}_k(i)\mathbf{x}_{k,Q}(i)e_{k,Q}^*(i)$

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

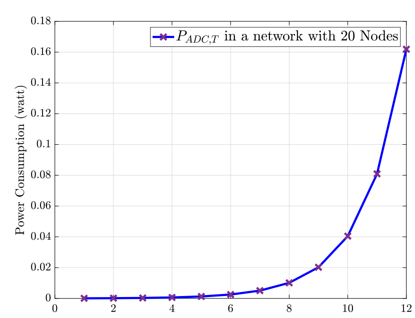
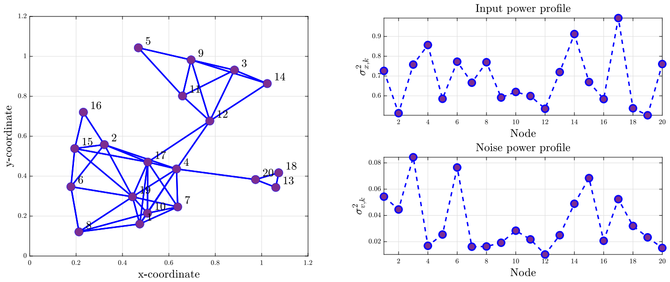


Fig. 2. Power consumption of the ADCs in an IoT network.

4. SIMULATION RESULTS

In this section, we assess the performance of the AdDQA-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



(a) Distributed network (b) Input and noise variances

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

generated by a white Gaussian noise process with variance $\sigma_{x,k}^2$ 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 $\sigma_{v,k}^2$. Fig. 3 plots the network details.

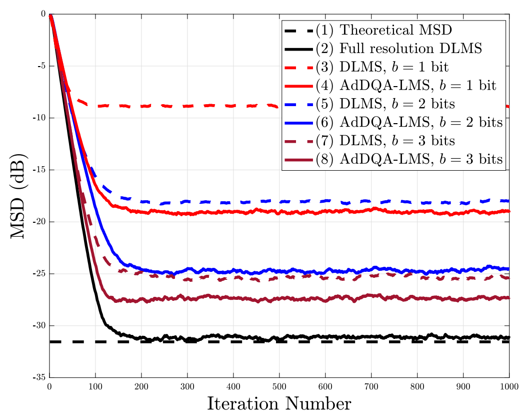


Fig. 4. MSD curves for DLMS and AdDQA-LMS algorithms.

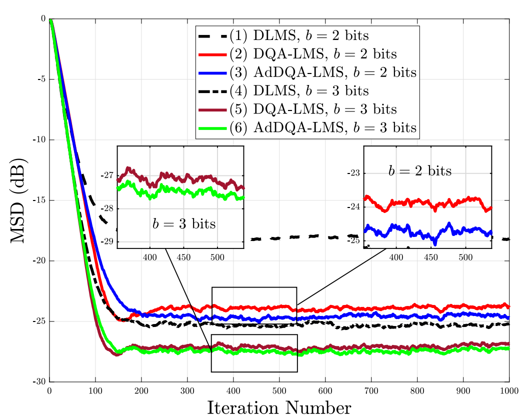


Fig. 5. The MSD curves for the DLMS, DQA-LMS, and AdDQA-LMS algorithms.

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 with different numbers of bits is assessed. The theoretical

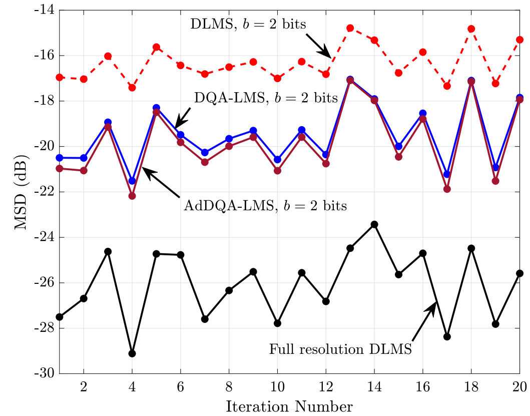


Fig. 6. The steady state MSD curves for the DLMS, DQA-LMS, and AdDQA-LMS algorithms.

MSD of the DLMS with the same step size μ and the Metropolis rule applied to a_{lk} is approximated by $\frac{\mu M}{N^2}\sum_{k=1}^{N}\sigma_{v,k}^2$ [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 AdDQA-LMS algorithm that improves the MSD for coarsely quantized signals. The performance of the proposed AdDQA-LMS algorithm is closer to the DLMS while it reduces about 90% of the power consumption of ADCs in the network (see Fig. 2).

In the next experiment, the MSD learning curves of the proposed AdDQA-LMS and DQA-LMS [12] are compared and the learning performance of the algorithms is shown in Fig. 5. It can be seen that the AdDQA-LMS improves the estimation performance of the DQA-LMS while both outperform the standard DLMS with low resolution signals coarsely quantized. Fig. 6 shows the node-wise steady state MSD values of the proposed AdDQA-LMS, DQA-LMS, and standard DLMS, by averaging the MSD values over the last 200 time samples.

5. CONCLUSION

In this paper, we have considered an energy-efficient framework for distributed learning and developed the AdDQA-LMS algorithm using low-resolution ADCs for adaptive IoT networks to improve the performance of the DQA-LMS. Simulations have shown the close performance of AdDQA-LMS to the DLMS algorithm while it enormously reduces the power consumption of the ADCs in the network operating with coarsely quantized signals.

6. REFERENCES

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

[2] M. M. Rana, W. Xiang, and E. Wang, "Iot-based state

estimation for microgrids," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1345–1346, 2018.

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

[4] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 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 Processing Magazine*, vol. 30, no. 3, pp. 155–171, 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 Processing*, 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 Processing Letters*, vol. 23, no. 4, pp. 551–555, 2016.

[8] S. Xu, R. C. de Lamare, and H. V. Poor, "Distributed compressed estimation based on compressive sensing," *IEEE Signal Processing Letters*, 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 *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 1556–1560.

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

[11] 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 Transactions on Signal Processing*, vol. 67, no. 6, pp. 1566–1581, 2019.

[12] A. Danaee, R. C. de Lamare, and V. H. Nascimento, "Energy-efficient distributed learning with coarsely quantized signals," *IEEE Signal Processing Letters*, vol. 28, pp. 329–333, 2021.

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

[14] A. Mezghani and J. A. Nossek, "Power efficiency in communication systems from a circuit perspective," in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*. IEEE, 2011, pp. 1896–1899.

[15] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on selected areas in communications*, vol. 17, no. 4, pp. 539–550, 1999.

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

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

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

[19] 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 Transactions on Wireless Communications*, vol. 17, no. 10, pp. 7080–7095, Oct 2018.

[20] 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 Communications Letters*, vol. 7, no. 3, pp. 476–479, June 2018.

[21] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," Tech. Rep. 216, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1952.

[22] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[23] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[24] 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 *2015 Information Theory and Applications Workshop (ITA)*. IEEE, 2015, pp. 191–198.

[25] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, "Nb-iot system for M2M communication," in *2016 IEEE wireless communications and networking conference*. IEEE, 2016, pp. 1–5.

[26] 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 65nm CMOS," in *2009 Symposium on VLSI Circuits*. IEEE, 2009, pp. 268–269.