Research Paper Presentation

Shambhu Prasad Kavir

CS20BTECH11045

Title and authors

Title

Energy Efficient Distributed Learning With Coarsely Quantized Signals

Authors

1. Alireza Danaee
2. Rodgrigo C. de Lamare
3. Vítor H. Nascimento

Introduction

- Distributed signal processing are of great relevance for statistical inference in wireless networks and applications such as wireless sensor networks and IoT. These techniques deal with the extraction of information from data collected at nodes that are distributed over a geographical area.
- 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 depends on the communication module, and from a circuit perspective on analog-to-digital converters(ADCs) and decoders.
- We present an energy-efficient distributed learning framework using low resolution ADCs and coarsely quantized signals for Internet of Things(IoT) networks.

Signal Model and Problem Statement

Consider an IoT network consisting of N nodes or agents which run distributed signal processing techniques to perform the desired tasks, as depicted in the figure:

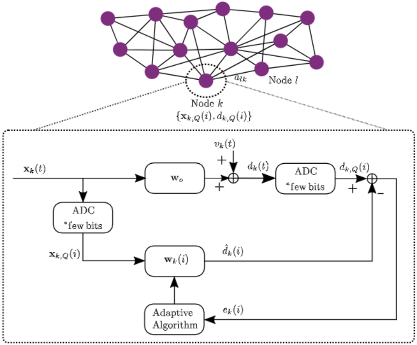


Figure: 1a) A distributed adaptive IoT network

Signal Model and Problem Statement

The model adopted considers a desired signal $d_k(i)$, at each time i, described by

$$d_k(i) = w_0^H x_k(i) + v_k(i) \qquad k = 1, 2, ..., N, \tag{1}$$

where $w_0 \in \mathbb{C}^{M\times 1}$ is the parameter vector that the agents must estimate, $x_k(i) \in \mathbb{C}^{M\times 1}$ is the regressor and $v_k(i)$ represents Gaussian noise with zero mean and variance $\sigma^2_{v,k}$ at node k.
We now adopt the Adapt-then-Continue diffusion rule. At each node k and time i, based on the local data $\{d_k(i), x_k(i)\}$ and the estimated parameter vectors $h_l(i)$ from its neighbourhood, the parameter vector with local estimates $w_k(i)$ is updated.

Signal Model and Problem Statement

The ATC distributed LMS(DLMS) algorithm consists of the recursions:

$$h_k(i) = w_k(i-1) = \mu_k x_k(i) e_k^*(i), \qquad w_k(i) = \sum_{l \in N_k} a_{lk} h_l(i), \tag{2}$$

- where $h_k(i)$ and $w_k(i)$ contain the intermediate and local estimates of w_0 at time i and node k, respectively,

$$e_k(i) = d_k(i) - \hat{d}_k(i) = d_k(i) - w_k^H(i-1)x_k(i) \tag{3}$$

- Here, $e_k(i)$ is the error between the output of the adaptive filter,$\hat{d}_k(i)$, and the the desired signal,$d_k(i)$, at time i,
- μ_k is the step-size for node k,
- N_k is the set of neighbour nodes connected to node k,

Signal Model and Problem Statement

- a_{lk} are the combination coefficents of the neighbour nodes at node k such that

$$a_{lk} = 0 \;\; if\, l \notin N_k, a_{lk} > 0 \;\; if\, l \in N_k, \;\; and \sum_{l \in N_k} a_{lk} = 1 \tag{4}$$

1. From figure 1a, as the measurement data in each node and the analog system are analog and each agent processes local data $\{d_k(i), x_k(i)\}$, we need two ADCs per agent.
2. As number of agents increases, the power consumption will grow considerably, when using high resolution ADCs for each agent.
3. This encourages us to quantize signals using fewer bits

Thus, the problem we are interested in solving is how to design energy-efficient learning algorithms that can cost effectively operate with coarsely quantized signals.

Proposed DQA-LMS Algorithm

- Let $x_{k,Q} = Q_b(x_k)$ 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 $\mathcal{L}_b = \{\ell_0, \ell_1, ...\ell_{2^b-1}\}$ where $\ell_p \in (\tau_p, \tau_{p+1}]$, for $p \in [0, 2^b - 1]$
- Let us assume $x_k \sim \mathcal{CN}(0, R_{x_k})$, where $R_{x_k} \in \mathbb{C}^{M \times M}$ is the covariance matrix of x_k.

Employing Bussgang's theorem, $x_{k,Q}$ can be decomposed as

$$x_{k,Q} = G_{k,b} x_k + q_k, \tag{5}$$

where the quantization distortion q_k is uncorrelated with x_k and $G_{k,b} \in \mathbb{R}^{M \times M}$ is a diagonal matrix described by

$$G_{k,b} = diag(R_{x_k})^{-\frac{1}{2}} \sum_{j=0}^{2^b-1} \frac{\ell_j}{\sqrt{\pi}} [\exp(-\mathcal{T}_j{}^2) diag(R_{x_k})^{-1}) \tag{6}$$

$$- \exp(-\mathcal{T}_{j+1}{}^2) diag(R_{x_k})^{-1})] \tag{7}$$

Proposed DQA-LMS Algorithm

- This signal decomposition is also applied to the desired signal $d_k(i)$, which is the output of the second ADC in the system, and for the particular case that $R_{x_k} = \mathbb{E}[x_k {x_k}^H] = \sigma^2_{x,k} I_M$, $G_{k,b}$ becomes $g_{k,b} I_M$.

To minimize the mean square error(MSE), between x_q and $x_{k,Q}$ we need to characterize the PDF of x_k to find the optimum quantization labels. Since the choice of labels based on PDF is not practical we assume the regressor x_k to be Gaussian, and approximate the labels and thresholds as follows:

- We generate an auxiliary Gaussian random variable with unit variance and then use Lloyd-Max algorithm to find a set of thresholds $\tilde{\mathcal{T}}_b = \{\tau_1, \tau_2, ..., \tau_{2^b-1}\}$ and labels $\tilde{\mathcal{L}}_b = \{\tilde{\ell}_0, \tilde{\ell}_1, ..., \tilde{\ell}_{2^b-1}\}$

Proposed DQA-LMS Algorithm

- We complete the set of thresholds $\mathcal{T}_b$ by adding $\tau_0 = -\infty$ and $\tau_{2^b} = \infty$ to the set $\tilde{\mathcal{T}}_b$
- 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 $\tilde{\mathcal{L}}_b$ by

$$\alpha = \left(2 \sum_{j=0}^{2^b-1} \tilde{\ell}_j^2 (\phi(\sqrt{2\mathcal{T}_{j+1}{}^2}) - \phi(\sqrt{2\mathcal{T}_j{}^2})) \right) \tag{8}$$

  to produce a set of suboptimal labels $\mathcal{L}_b = \alpha \tilde{\mathcal{L}}_b$, where ϕ is the CDF of a standard Gaussian random variable.

We generate these thresholds and labels offline to build $G_{k,b}$ for the proposed DQA-LMS algorithm

Derivation of DQA-LMS

- We consider $x_k(t)$ and $d_k(t)$ as the analog input and output of the unknown system w_0 at node k. Let $x_k(i)$ and $d_k(i)$ denote the high precision sampled versions of $x_k(t)$ and $d_k(t)$, and $x_{k,Q}(i)$ and $d_{k,Q}(i)$ denote the coarsely quantized versions of $x_k(i)$ and $d_k(i)$ respectively.
- We assume the input signal at each node is Gaussian with zero mean and covariance matrix $R_{x_k} = \mathbb{E}[x_k {x_k}^H] = \sigma^2_{x,k} I_M$ for $k = 1, 2, ..., N$. Using equation (5) we can decompose $x_{k,Q}(i)$ and $d_{k,Q}(i)$ as

$$x_{k,Q}(i) = g_{k,b}(i) x_k(i) + q_{x,k}(i) \tag{9}$$

$$d_{k,Q}(i) = Q(d_k(i) \approx g_{k,b}(i) d_k(i) + q_k(i) \tag{10}$$

$$= g_{k,b}(i) w_0^H x_k(i) + \hat{q}_k(i) \tag{11}$$

Derivation of DQA-LMS

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

Derivation of DQA-LMS

- We show next that a learning algorithm based directly on (10) is biased for estimating w_0, and show how to correct for this bias.
- For this, let $\beta_k(i)$ be a coefficient to be chosen shortly, and define $\hat{d}_k(i) = \beta_k(i) w_k^H(i-1) x_{k,Q}(i)$ and construct an MSE cost function defined by,

$$\begin{aligned} J_k(w_k(i)) &= \mathbb{E}[|e_{k,Q}(i)|^2] & (12) \\ &= \mathbb{E}[|d_{k,Q}(i) - \hat{d}_k(i)|^2] & (13) \\ &= \mathbb{E}[|d_{k,Q}(i) - \beta_k(i) w_k^H(i-1) x_{k,Q}(i)|^2] & (14) \end{aligned}$$

which depends only on the observed quantized quantities $x_{k,Q}(i)$ $d_{k,Q}(i)$. For $\beta_k(i) = 1$ as in DLMS, the quantization of $d_k(i)$ would result in biased estimated of w_0.

Derivation of DQA-LMS

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 (14) is described by

$$h_k(i) = w_k(i-1) - \mu_k \nabla J_k(w_k(-1)) \tag{15}$$

To compute the gradient of (15) we write the error in (14) as

$$\begin{aligned} e_{k,Q}(i) &= d_{k,Q}(i) - \beta_k(i) w_k^H(i-1) x_{k,Q}(i) && (16) \\ &= g_{k,b}(i) w_0^H x_k(i) + \hat{q}_k(i) - \beta_k(i) && (17) \\ & w_k^H(i-1)(g_{k,b}(i) x_k(i) + q_{x,k}(i)) && (18) \\ &= g_{k,b}(i)(w_0^H(i) - \beta_k(i) w_k^H(i-1) x_k(i)) && (19) \\ &- \beta_k(i) w_k^H(i-1) q_{x,k}(i) + \hat{q}_k(i) && (20) \end{aligned}$$

Derivation of DQA-LMS

We assume that $R_{x_k} = \mathbb{E}[x_k x_k^H] = \sigma_{x,k}^2 I_M$ and $R_{q,k} = \mathbb{E}[q_{x,k} q_{x,k}^H = \sigma^2 q, k I^M$ .Substituting (20) in (14) and taking the expected value of (15), we have

$$\mathbb{E}[h_k(i)] = [I_M - \mu_k g_{k,b}^2(i)\beta_k(i)R_{x_k}] \tag{21}$$

$$\mathbb{E}[w_k(i-1)] + \mu g_{k,b}^2(i)R_{x_k} w_0^H \tag{22}$$

Substituting the values of R_{x_k} and $R_{q,k}$ and taking the limit on (21), we obtain

$$\lim_{i \to +\infty} \mathbb{E}[h_k(i)] = \frac{1}{\beta_k(i)} \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2} \tag{23}$$

We conclude that the solution is unbiased if we choose,

$$\beta_k(i) = \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2} \tag{24}$$

Derivation of DQA-LMS

The gradient of $|e_{k,Q}(i)|^2$ with respect to w_0^H is $\nabla J_k(w_k(i-1)) = -\frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2+\sigma_{q,k}^2} x_{k,Q}(i)e_{k,Q}^*(i)$. After organising the terms of the gradient, we obtain the DQA-LMS algorithm:

$$h_k(i) = w_k(i-1) + \frac{1}{\beta_k(i)} \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2} x_{k,Q}(i)e_{k,Q}^*(i) \tag{25}$$

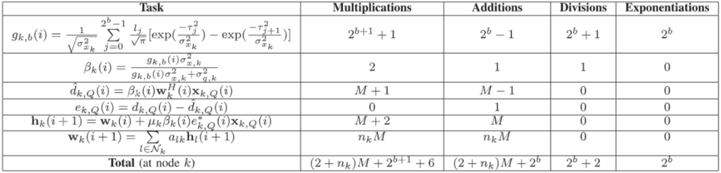
$$w_k(i) = \sum_{l \in \mathcal{N}_k} a_{lhk} h_l(i) \tag{26}$$

$$e_{k,Q}(i) = d_{k,Q}(i) - \frac{g_{k,b}(i)\sigma_{x,k}^2}{g_{k,b}(i)\sigma_{x,k}^2 + \sigma_{q,k}^2} w_k^H(i-1)x_{k,Q}(i) \tag{27}$$

$$g_{k,b}(i) = \frac{1}{\sqrt{\sigma_{x_k}^2}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} [\exp(\frac{-\tau_j^2}{\sigma_{x_k}^2}) - \exp(\frac{-\tau_{j+1}^2}{\sigma_{x_k}^2})] \tag{28}$$

Computational Complexity and Energy Consumption

The following table shows the computational complexity of the DQA-LMS algorithm in terms of Number of additions and multiplications at node k per time instant, where n_k is the number of neighbour nodes connected to node k.



Task	Multiplications	Additions	Divisions	Exponentiations
$g_{k,b}(i) = \frac{1}{\sqrt{\sigma^2_{x_k}}} \sum_{j=0}^{2^b-1} \frac{l_j}{\sqrt{\pi}} [\exp(\frac{-\tau_j^2}{\sigma^2_{x_k}}) - \exp(\frac{-\tau_{j+1}^2}{\sigma^2_{x_k}})]$	$2^{b+1}+1$	2^b-1	2^b+1	2^b
$\beta_k(i) = \frac{g_{k,b}(i)\sigma^2_{x,k}}{g_{k,b}(i)\sigma^2_{x,k}+\sigma^2_{q,k}}$	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

Figure: 1b) Computational Complexity Per Time Instant

- At each time instant,the DQA-LMS performs a few more operations($\approx O(2^b)$) than DLMS.
- However, the extra complexity of DQA-LMS allows the system to work in a more energy-efficient way.

Computational Complexity and Energy Consumption

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

$$P_{ADC}(b) = cB2^b \tag{29}$$

Here,

- B is the bandwith(related to sampling rate),
- b is the number of quantization bits of the ADC,
- c is the power consumption per conversion step.

Therefore, the total power consumption of the ADCs in the network is

$$P_{ADC,T}(b) = 2NcB2^b \tag{30}$$

Computational Complexity and Energy Consumption

The following figure shows an example of the total power consumption of the ADCs in a narrowband IoT(NB-IoT) network running diffusion adaptation consisting of 20 nodes with bandwidth $B = 200kHz$ and considering the power consumption per conversion step of each ADC, $c = 494fJ$

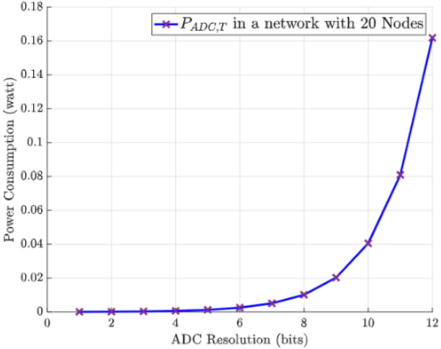
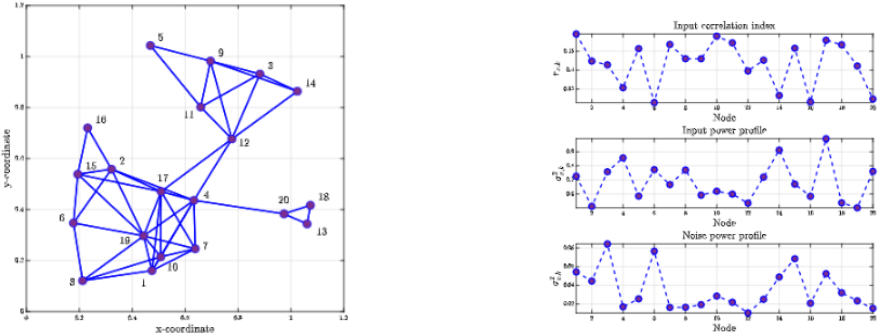


Figure: 1c) Power consumption of ADCs in an adaptive IoT network

Simulation Results

- Here, 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 system of the unknown system has $M = 8$ taps, is generated normally and normalised to one.
- The input signals $x_k(i)$ at each node are generated by passing a white Gaussian noise process with $\sigma^2_{x,k}$ through a first order regessive 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 $x_{k,Q}(i)$.
- The noise samples from each node are drawn from a zero mean white Gaussian process with variance $\sigma^2_{v,k}$.

Simulation Results



(a) Distributed network structure

(b) Variance and correlation coefficients

Figure: A wireless network with $N = 20$ nodes

Simulation Results

The simulated mean square deviation(MSD) learning curves are obtained by ensemble averaging over 100 independent trials.We choose the same step size for all agents, i.e., $\mu_k = 0.05$. The combining coefficients a_{lk} are computed by the Metropolis rule.

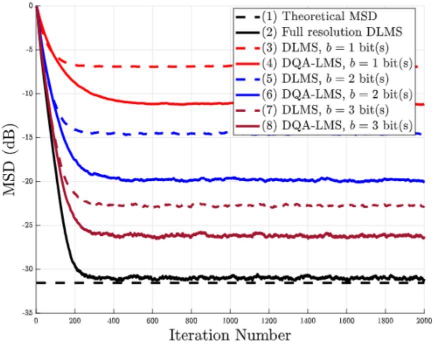


Figure: 1E) The MSD curves for DLMS and DQA-LMS algorithms

Simulation Results

- Curve 1 shows the theoretical MSD of DLMS.
- 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 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 propsoed 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 the ADCs in the network.