Citation/Reference	Authors: Matthias Blochberger, Jan Østergaard, Randall Ali, Marc Moonen, Filip Elvander, Jesper Jensen, Toon van Waterschoot Title: Adaptive Coding in Wireless Acoustic Sensor Networks for Distributed Blind System Identification Conference: 2023 Asilomar Conference on Signals, Systems, and Computers
Dates	**Conference paper** • Date of conference: October 29th - November 2nd, 2023
Archived version	☐ Submitted version ■ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ■ FWO ☐ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Author contact	matthias.blochberger@esat.kuleuven.be

ADAPTIVE CODING IN WIRELESS ACOUSTIC SENSOR NETWORKS FOR DISTRIBUTED BLIND SYSTEM IDENTIFICATION

M. Blochberger [1*], *J. Østergaard* [2], *R. Ali* [3], *M. Moonen* [1], *F. Elvander* [4], *J. Jensen* [2], *T. van Waterschoot* [1]

[1] KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), Leuven, Belgium
[2] Aalborg University, Dept. of Electronic Systems, Aalborg, Denmark
[3] University of Surrey, Dept. of Music & Media, Guildford, UK
[4] Aalto University, Dept. of Information and Communications Engineering, Espoo, Finland
[*] *matthias.blochberger@esat.kuleuven.be*

ABSTRACT

With distributed signal processing gaining traction in the audio and speech processing landscape through the utilization of interconnected devices constituting wireless acoustic sensor networks, additional challenges arise, including optimal data transmission between devices. In this paper, we extend an adaptive distributed blind system identification algorithm by introducing a residual-based adaptive coding scheme to minimize communication costs within the network. We introduce a coding scheme that takes advantage of the convergence of estimates, i.e., vanishing residuals, to minimize information being sent. The scheme is adaptive, i.e., tracks changes in the estimated system and utilizes entropy coding and adaptive gain to fit the time-varying residual variance to pre-trained codebooks. We use a low-complexity approach for gain adaptation, based on a recursive variance estimate. We demonstrate the approach's effectiveness with numerical simulations and its performance in various scenarios.

Index Terms— adaptive coding, blind system identification, alternating direction method of multipliers

1. INTRODUCTION

Blind system identification is an acoustic signal-processing task that focuses on identifying the characteristics of an unknown acoustic system using only its output signals. In other words, the goal is to estimate a model of the system, such as its impulse or frequency response, without knowledge of the input signal of the system. Various techniques have been developed to estimate impulse and frequency responses of the system in question, such as [1, 2]. We consider single-input-multiple-output (SIMO) systems, i.e., one acoustic source and multiple acoustic sensors in a room.

Tackling the problem of estimating impulse responses within the wireless acoustic sensor networks (WASN) [3], i.e., a set of sensor nodes with processing capabilities that can communicate with each other gives rise to challenges regarding the inter-node communication, such as limited transmission bandwidth, which we address in this paper. In previous work [4], we proposed an adaptive distributed algorithm for estimating impulse responses using the general-form consensus alternating direction method of multipliers (ADMM) [5, 6], avoiding the necessity of a fusion center for multi-channel signal processing. Moreover, relevant existing literature deals with quantisation in distributed algorithms [7, 8], adaptive entropy encoding [9] and also learning-based quantisation in ADMM [10, 11] as well as the effects of quantisation in PDMM [12].

We extend the existing algorithm [4] to use residual transmissions, which we assume have approximately normal distributions. Using this assumption, we introduce an adaptive coding scheme that utilizes Huffman codebooks [13]. Instead of adapting the encoder and decoder directly to the time-varying distribution of residuals, we introduce a simple approach where the residuals are scaled before encoding and inversely descaled after decoding, effectively resulting in a gain-shape approach [14]. This simplifies the method greatly, as it avoids adapting codebooks which may be computationally inefficient. To adapt the gain, we propose a low-complexity approach using a straightforward time-recursive estimate of the decoded residual's variances, which is compared to the variance of the data used for codebook training, and the data scaled accordingly. Finally, we use numerical simulations to demonstrate the method's effectiveness.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the SOUNDS European Training Network. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369. This research received funding in part from the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (No. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. The simulation code for this work can be found at https://github.com/SOUNDS-RESEARCH/asilomar2023-adaptive-coding.

2. PROBLEM STATEMENT AND BSI IN WASN

The WASN consists of M nodes indexed by $i \in \mathcal{M} \triangleq \{1, ..., M\}$ and the directed edges $(i, j) \in \mathcal{E}$, i.e., connections for information exchange. We denote the node neighborhoods defined by these edges as $\mathcal{N}_i \subseteq \mathcal{M}$, and $N_i = |\mathcal{N}_i|$ is the size of the neighborhood. We treat this problem under the following assumptions:

- Network topology $\mathcal{G} = (\mathcal{M}, \mathcal{E})$ is static
- Network topology is connected (there exists an (indirect) path from each node $i \in \mathcal{M}$ to each other node $j \in \mathcal{M}$)
- Communication is instantaneous/synchronous
- Communication is error-free apart from losses introduced by the coding scheme

We consider a single-input-multiple-output (SIMO) acoustic system represented by the stacked vector

$$\mathbf{h} = \left[\left\{\mathbf{h}_i^{\mathrm{T}}\right\}_{i\in\mathcal{M}}\right]^{\mathrm{T}}, \tag{1}$$

where $\mathbf{h}_i$, $i \in \mathcal{M}$, are $L \times 1$ impulse responses of the system. The signal model

$$\mathbf{x}_i(k) = \mathbf{H}_i\mathbf{s}(k) + \mathbf{v}_i(k), \tag{2}$$

where $\mathbf{H}_i$ is the $2L \times 2L$ diagonal filtering matrix of the ith channel using the $2L$-DFT transform of $\mathbf{h}_i$, zero-padded to $2L \times 1$. The vectors $\mathbf{s}(k)$ and $\mathbf{v}_i(k)$ are the DFT transforms of the zero-padded source signal and additive white Gaussian noise at frame k respectively. The distributed cross-relation minimization problem (cf. [4]) with regularization is

$$\min_{\{\mathbf{w}_i, \mathbf{h}_i\}} \quad \sum_{i\in\mathcal{M}} f_i(\mathbf{w}_i) \tag{3a}$$

$$\text{s.t.} \quad \mathbf{w}_{i|j} = \mathbf{z}_{j|j} \quad i \in \mathcal{M}, j \in \mathcal{N}_i \tag{3b}$$

$$\|\mathbf{z}\| = 1 \quad i \in \mathcal{M} \tag{3c}$$

where

$$f_i(\mathbf{w}_i) = \frac{1}{2}\mathbf{w}_i^{\mathrm{H}}\mathbf{P}_i\mathbf{w}_i + \frac{\lambda}{2}\|\mathbf{w}_i\|^2 \tag{4}$$

is the regularized cross-relation cost function. The regularization term ensures small values for $\mathbf{w}_i$, which is beneficial to the algorithm's stability. The subscript $_{i|j}$ denotes a subvector of a stacked vector. We define this for $\mathbf{w}_i$, which is an $N_iL \times 1$ stacked vector of $L \times 1$ subvectors $\mathbf{w}_{i|j}$ of the form

$$\mathbf{w}_i = \left[\left\{\mathbf{w}_{i|j}^{\mathrm{T}}\right\}_{j\in\mathcal{N}_i}\right]^{\mathrm{T}} \tag{5}$$

where $\mathbf{w}_{i|j}$ is a node-local estimate of $\mathbf{h}_j$ at node i. The variables $\mathbf{z}_i$ and $\mathbf{z}_{i|j}$, the neighborhood-consensus estimate of $\mathbf{h}_j$ at node i, are defined analogously. The scalar $0 < \lambda \ll 1$ is the regularisation parameter. The recursively updated matrix $\mathbf{P}_i$ is the cross-relation matrix [2, 4], a block matrix consisting of cross-spectral density matrices of pairs of channel signals $(\mathbf{x}_i(k), \mathbf{x}_j(k))$ to form a system of linear equations. The general-form consensus constraint (3b) forces the convergence to a global solution for parameters estimated by multiple nodes, while the non-triviality constraint (3c) is necessary to avoid convergence to the zero vector.

We formulate the augmented Lagrangian for (3) as the real-valued function of complex variables,

$$\mathcal{L}_\rho(\mathbf{w}_1, ..., \mathbf{w}_M, \mathbf{z}_1, ..., \mathbf{z}_M, \mathbf{u}_1, ..., \mathbf{u}_M) = \sum_{i\in\mathcal{M}} \left(f_i(\mathbf{w}_i) + 2\Re\left(\mathbf{u}_i^{\mathrm{H}}(\mathbf{w}_i - \mathbf{z}_i)\right) + \frac{\rho}{2}\|\mathbf{w}_i - \mathbf{z}_i\|^2\right) \tag{6}$$

where $\mathbf{u}_i$ is defined analogously to (5) with the subvectors $\mathbf{u}_{i|j}$, the Lagrange multipliers. The scalar $\rho > 0$ is the ADMM penalty parameter, and $\mathbf{u}_i$ is the dual variable - Lagrangian multiplier - of the consensus constraint [5], defined analogously to (5). The update steps for primal and dual variables can be readily verified as

$$\mathbf{w}_i^{(k+1)} = \mathbf{w}_i^{(k)} - \mu\left(\Delta_{\mathbf{w}_i^*}\mathcal{L}_\rho\right)^{-1}\nabla_{\mathbf{w}_i^*}\mathcal{L}_\rho \tag{7a}$$

$$\mathbf{y}_{i|j}^{(k+1)} = \rho\,\mathbf{w}_{i|j}^{(k+1)} + \mathbf{u}_{i|j}^{(k)} \tag{7b}$$

$$\mathbf{y}_{j|i}^{(k+1)} = \mathbf{y}_{i|j}^{(k+1)} \tag{7c}$$

$$\mathbf{z}_{i|i}^{(k+1)} = \mathcal{P}_{\|\mathbf{z}\|=1}\left(\frac{1}{N_i}\sum_{j\in\mathcal{N}_i}\mathbf{y}_{i|j}^{(k+1)}\right) \tag{7d}$$

$$\mathbf{z}_{j|i}^{(k+1)} = \mathbf{z}_{i|i}^{(k+1)} \tag{7e}$$

$$\mathbf{u}_i^{(k+1)} = \mathbf{u}_i^{(k)} + \rho\left(\mathbf{w}_i^{(k+1)} - \mathbf{z}_i^{(k+1)}\right) \tag{7f}$$

where $\nabla_{\mathbf{w}_i^*}\mathcal{L}_\rho$ is the gradient, $\Delta_{\mathbf{w}_i^*}\mathcal{L}_\rho$ the Hessian with regards to $\mathbf{w}_i^*$ of the augmented Lagrangian (6), $\mathcal{P}_{\|\mathbf{z}\|=1}$ is an operator enforcing (3c) by, e.g., normalization and μ is a step size. The local combination of estimate and dual variable, $\mathbf{y}_i$ is defined analogously to (5), with subvectors $\mathbf{y}_{i|j}$. The equations (7c) and (7e) denote copying variables along edges $(i, j) \in \mathcal{E}$ of the network, which represents a transmission between nodes in practical terms. In the following, we aim to find a low-complexity coding scheme that reduces the amount of data transmitted within the WASN while minimizing the effect on the performance of the estimation algorithm.

3. PROPOSED METHOD

The total bitrate within the network for uncompressed transmissions of the variables $\mathbf{y}_{i|j}^{(k+1)}$ and $\mathbf{z}_{i|j}^{(k+1)}$ is constant at $R^{(k)} = \sum_{i\in\mathcal{M}} 2N_iLB$, where B is the number of bits used for the representation of the vector elements, e.g., 128 for complex values on standard architectures. For large vectors and networks, this can become significantly power-intensive, therefore the goal is its reduction.

For brevity and readability, and since the method is analogous for either $\mathbf{y}_{j|i}^{(k)}$ and $\mathbf{z}_{j|i}^{(k)}$, let us use a placeholder variable $\mathbf{d}^{(k)}$ - for data. As in [12], instead of transmitting a sequence of variables $\{\mathbf{d}^{(k)}\}_{k=1}^{K}$, we transmit a sequence of residuals $\{\mathbf{r}^{(k)}\}_{k=1}^{K}$, i.e., the difference vectors

$$\mathbf{r}^{(k)} = \mathbf{d}^{(k)} - \tilde{\mathbf{d}}^{(k-1)}, \tag{8}$$

where $\tilde{\mathbf{d}}^{(k-1)}$ is the reconstructed data after encoding, transmission and decoding. This reconstruction is defined as

$$\tilde{\mathbf{d}}^{(k)} = \tilde{\mathbf{r}}^{(k)} + \tilde{\mathbf{d}}^{(k-1)} \tag{9}$$

and combines the preceding iteration's reconstruction $\tilde{\mathbf{d}}^{(k-1)}$ and the current iteration's encoded-decoded difference vector $\tilde{\mathbf{r}}^{(k)}$. Note that the reconstructed quantities $\tilde{\mathbf{r}}^{(k)}$ and $\tilde{\mathbf{d}}^{(k)}$ have to be computed at both transmitting and receiving nodes, and the latter stored to be available for residual computation at the next iteration (8). We define encoder ϕ and decoder ψ for a residual sequence $\{\mathbf{r}^{(k)}\}_{k=1}^{K}$ such that

$$\{\tilde{\mathbf{r}}^{(k)}\}_{k=1}^{K} = \psi\left(\phi\left(Q\left(\{\mathbf{r}^{(k)}\}_{k=1}^{K}\right)\right)\right). \tag{10}$$

The operator Q is the element-wise projection of the residual sequences $\{\mathbf{r}^{(k)}\}_{k=1}^{K}$ real and imaginary parts onto the set of values represented by the codebook $\mathcal{R}^{K}$. We define (element-wise) encoder and decoder as the mappings

$$\phi : \mathcal{R}^{K} \to \{1, ..., N\} \tag{11a}$$
$$\psi : \{1, ..., N\} \to \mathcal{R}^{K}, \tag{11b}$$

where the integers $\{1, \ldots, N\}$ are symbols sent as information.

We employ entropy coding to reduce the number of bits per transmission. For the encoder and decoder, we make use of the assumption that the residual values are approximately normal-distributed, which is supported by numerical tests with an input signal drawn from a normal distribution, see Fig. 1 (left). To generate entropy codebooks, a specified number of symbols in $\mathcal{R}^{K}$ is used with their corresponding probability values according to a normal distribution. Fig. 1 (right) illustrates two examples of possible sets $\mathcal{R}^{K}$ with $N = |\mathcal{R}^{K}| = 5$ and $N = 13$ and their probabilities. We impose the constraint that there has to be a symbol for residual value 0. These sets of value-probability pairs are used to generate Huffman codebooks [13], which yield binary representations of symbols (11), where more probable symbols are generally represented using fewer bits than less probable ones. Additionally, reducing the number of symbols limits the maximum number of bits; therefore we want to use as few as possible while avoiding impact on algorithm performance.

To generate such a codebook, the variance of the residual's normal distribution has to be known. However, this

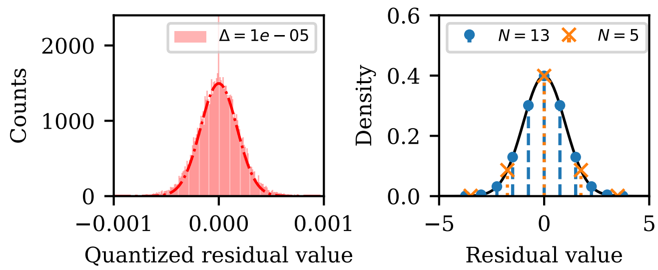

Fig. 1: Example distribution of residual values and overlayed normal distribution.

is highly dependent on scenario and algorithm parameters, such as acoustic SNR and especially the convergence state of the estimation algorithm. Therefore, an adaptive solution is preferable. The approach in [15], a pre-defined geometric contraction of the quantisation interval with 1-bit quantisation is not applicable as it requires fine-tuning of parameters according to the problem's convergence speed and requires a static fixed point, i.e., a static optimal solution. Neither can be said in this case, however.

One approach would be to train many codebooks on a set of expected variances $\sigma^2 \in \left[\sigma^2_{\text{min}}, \sigma^2_{\text{max}}\right]$ and switch the codebook used by the encoder and decoder, $\psi^{(k)}, \phi^{(k)}$, accordingly at each or at certain frames (k). However, since we assume that the residuals are normal-distributed, regardless of variance, we can simplify this approach by using a single-codebook scheme ϕ_0 , ψ_0, with gain factor $\delta^{(k)}$ such that

$$\frac{1}{\delta^{(k)}}\psi_0\left(\phi_0\left(\delta^{(k)}x^{(k)}\right)\right) \approx \psi^{(k)}\left(\phi^{(k)}\left(x^{(k)}\right)\right), \tag{12}$$

in which the approximation is due to nonlinearities in the encoder and decoder. The codebook $\mathcal{R}_0^{K}$ is generated with N symbols and probabilities according to

$$p(r) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2}\frac{r^2}{\sigma_0^2}\right). \tag{13}$$

This yields a simple gain-shape approach [14] that scales the data to fit the encoder rather than vice-versa. For the appropriate choice of $\delta^{(k)}$, we introduce a simple method.

Under the assumption of a zero-mean distribution, we estimate the variance of the decoded residual $\tilde{\mathbf{r}}^{(k)}$ by the recursion [16]

$$(\hat{\sigma}^2_{\tilde{\mathbf{r}}})^{(k)} = \alpha(\hat{\sigma}^2_{\tilde{\mathbf{r}}})^{(k-1)} + (1-\alpha)\frac{1}{L}\|\tilde{\mathbf{r}}^{(k)}\|^2 \tag{14}$$

where $(\hat{\sigma}^2_{\tilde{\mathbf{r}}})^{(k)}$ is the estimate at time index k, $\|\tilde{\mathbf{r}}^{(k)}\|$ is the Euclidian norm of the reconstructed residual vector (9) and $0 < \alpha < 1$ is an exponential smoothing factor. This yields a scalar variance estimate for the vector $\tilde{\mathbf{r}}^{(k)}$, preceded by the assumption that all vector elements have approximately similar distributions. Note that this quantity can be estimated at the encoder and decoder independently and - assuming error-free transmission - the estimates are equivalent. Here, the

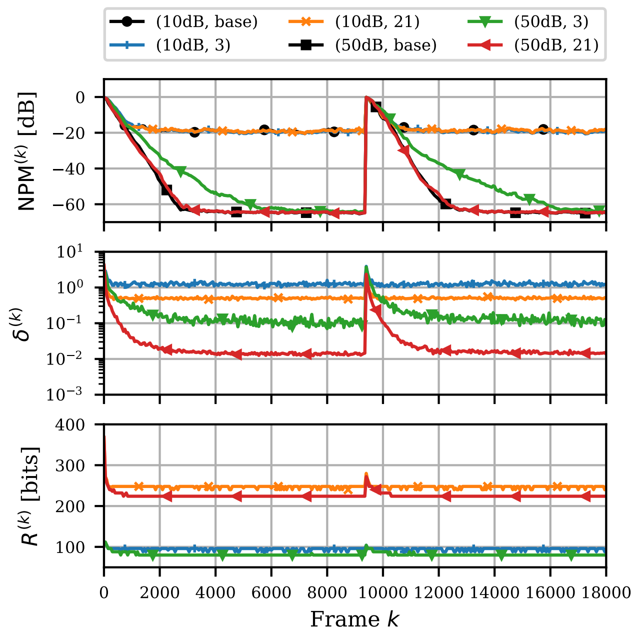

Fig. 2: **(Top)** NPM is plotted over frames to show convergence behaviour. **(Middle)** The gain factor $\delta^{(k)}$ over frames. **(Bottom)** Rate $R^{(k)}$ over frames. The legend denotes the parameter combinations (SNR, N) for adaptive approach or $(\text{SNR}, \text{"}base\text{"})$ for the baseline. For all, medians of 100 Monte Carlo realisations are shown.

gain factor $\delta^{(k)}$ follows trivially from estimated and training variances as

$$\delta^{(k)} = \left[\frac{\sigma_0^2}{(\hat{\sigma}_{\tilde{\mathbf{r}}}^2)^{(k-1)}} \right]^p \tag{15}$$

where $p = 0.5$. In practice, it was observed that using a $p > 0.5$ improves the results for certain scenarios, as it likely compensates for overestimating the variance.

4. NUMERICAL SIMULATION

We conducted a numerical simulation study to show the proposed method's effectiveness compared to a baseline approach. This baseline uses a 128-bit complex floating-point representation for each element of the residuals which are assumed to be transmitted as such. We consider the following measures for evaluation: The normalized projection misalignment [17] between the true system impulse responses $\mathbf{h}$ and the consensus estimate $\mathbf{z}^{(k)}$,

$$\text{NPM}^{(k)} = 20 \log_{10} \left(\left\| \mathbf{h} - \frac{\mathbf{h}^{\mathrm{T}} \mathbf{z}^{(k)}}{\mathbf{h}^{\mathrm{T}} \mathbf{h}} \mathbf{z}^{(k)} \right\| / \left\| \mathbf{z}^{(k)} \right\| \right).$$

The rate $R^{(k)}$ within the WASN is the total number of bits exchanged between all nodes per frame.

For the evaluation, we simulated a sensor network with $M = 3$ nodes and directed edges $\mathcal{E} = \{(1,2), (2,3), (3,1)\}$ under the following scenarios, of which the parameter values are listed in Table 1: The input signal is white, and i.i.d normal-distributed and N_{signal} samples long. The impulse responses of length L are drawn from a normal distribution as well. Further, the signal-to-noise ratio (SNR), where additive white noise is uncorrelated with the input signal and independent across channels, and the number of symbols N are varied over a set of values. The codebook $\mathcal{R}_0^K$ is generated with variance σ_0^2. All M impulse responses are changed at $\frac{1}{2} N_{\text{signal}}$ to test the algorithm's ability to adapt to a time-varying system. The parameters ρ, μ, λ of the ADMM-BSI algorithm as well as coding parameters α, p are kept constant over all combinations of SNR and N and all 100 Monte-Carlo simulation runs.

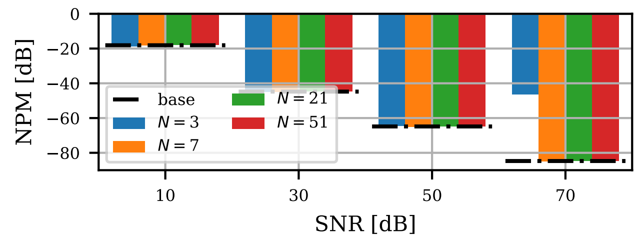

Fig. 3: Average NPM at selected time frames $k \in \{8500, 18000\}$.

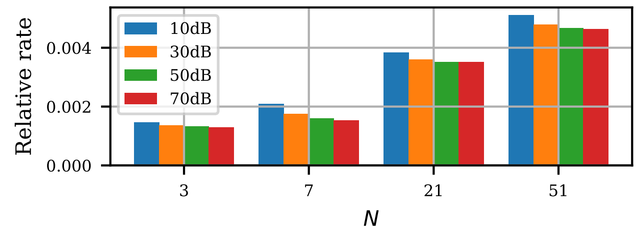

Fig. 4: Time-average of relative rate $R^{(k)}/R^{(k)}_{\text{base}}$, for N and SNR as listed in Table 1.

Fig. 2 shows examples of convergence behaviour at two values of SNR and the number of symbols $N \in \{3, 21\}$ compared to the baseline approach. It shows that the coding scheme manages to avoid a big impact on estimation performance. The steady-state estimation error measured by NPM after convergence is not affected; however, some impact on convergence speed is observable at low N and high SNR. This can be observed in Fig. 3 as well, where at low N and high SNR the algorithm has not converged at the selected time frames.

It further demonstrates adaptivity by reconverging after abruptly changing the impulse responses. Most importantly, the number of bits used per iteration, the rate $R^{(k)}$, is significantly reduced, see Fig. 2 (bottom) and for its time-averages Fig. 4.

P	Eq	V	P	Eq	V
MC	–	100	λ	(7b)	0.01
SNR	–	$\{10, 30, 50, 70\}$	ρ	(7b) (7f)	1.0
N_{signal}	–	3×10^5	α	(14)	0.6
L	(14)	16	p	(15)	2.0
N	(11)	$\{3, 7, 21, 51\}$	σ_0^2	(13) (15)	10^{-4}
μ	(7a)	0.8			

Table 1: Simulation parameters. P = Parameter, Eq = Equation, V = Value. MC = Monte-Carlo realisations.

5. CONCLUSION

In this paper, we introduce an adaptive coding scheme to our previous work on distributed blind system identification, i.e., impulse response estimation, in wireless acoustic sensor networks. We use entropy coding, specifically Huffman codebooks, to reduce the number of bits per transmission. We describe the method to compute a gain factor, which is used to fit the residual data to be transmitted to the encoder-decoder's useable value range. This is based on a recursively estimated variance of the decoded residuals. We demonstrate the efficacy of the coding scheme in numerical simulations. An extension to this could be the investigation of the influence of the structure of impulse responses on the behaviour of residuals and vector encoding.

6. REFERENCES

[1] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEICE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[2] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[3] C. Yu, L. Xie, and Y. C. Soh, "Distributed blind system identification in sensor networks," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '14)*, May 2014, pp. 5065–5069.

[4] M. Blochberger, F. Elvander, R. Ali, M. Moonen, J. Østergaard, J. Jensen, and T. van Waterschoot, "Distributed cross-relation-based frequency-domain blind system identification using Online-ADMM," in *Proc. 2022 Int. Workshop Acoustic Signal Enhancement (IWAENC '22)*, 2022, pp. 1–5.

[5] S. Boyd, N. Parikh, E. Chu, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[6] Q. Ling and A. Ribeiro, "Decentralized dynamic optimization through the alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1185–1197, 2014.

[7] A. Reisizadeh, A. Mokhtari, H. Hassani, and R. Pedarsani, "Quantized decentralized consensus optimization," in *Proc. 57th IEEE Conf. Decision Control (CDC '18)*, 2018, pp. 5838–5843.

[8] S. Zhu and B. Chen, "Quantized consensus by the admm: Probabilistic versus deterministic quantizers," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1700–1713, 2016.

[9] D. L. Duttweiler and C. Chamzas, "Probability estimation in arithmetic and adaptive-huffman entropy coders," *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 237–246, 1995.

[10] Y. Liu, K. Yuan, G. Wu, Z. Tian, and Q. Ling, "Decentralized dynamic admm with quantized and censored communications," in *Conf. Rec. 53rd Asilomar Conf. Signals, Syst. Computers*, 2019, pp. 1496–1500.

[11] T. X. Nghiem, A. Duarte, and S. Wei, "Learning-based adaptive quantization for communication-efficient distributed optimization with admm," in *Conf. Rec. 54th Asilomar Conf. Signals, Syst. Computers*, 2020, pp. 37–41.

[12] D. H. M. Schellekens, T. Sherson, and R. Heusdens, "Quantisation effects in pdmm: A first study for synchronous distributed averaging," in *Proc. 2017 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '17)*, 2017, pp. 4237–4241.

[13] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[14] R. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, no. 2, pp. 4–29, 1984.

[15] J. A.G. Jonkman, T. Sherson, and R. Heusdens, "Quantisation effects in distributed optimisation," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '18)*, 2018, pp. 3649–3653.

[16] M. M. Bruce, *Estimation of variance by a recursive equation*, vol. 5465, National Aeronautics and Space Administration, 1969.

[17] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, 1998.