Block-sparsity-aware LMS algorithm for network echo cancellation

Wei, Ye; Zhang, Yonggang; Wang, Chengcheng

2018

Wei, Y., Zhang, Y., & Wang, C. (2018). Block-sparsity-aware LMS algorithm for network echo cancellation. *Electronics Letters*, 54(15), 951-953. doi:10.1049/el.2018.1065

<https://hdl.handle.net/10356/89501>

<https://doi.org/10.1049/el.2018.1065>

© 2018 The Institution of Engineering and Technology. This paper was published in *Electronics Letters* and is made available as an electronic reprint (preprint) with permission of The Institution of Engineering and Technology. The published version is available at: [<http://dx.doi.org/10.1049/el.2018.1065>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 14:21:52 SGT

Block-sparsity-aware LMS algorithm for network echo cancellation

Ye Wei, Yonggang Zhang[✉] and Chengcheng Wang

Network echo path impulse response is single-block-sparse in nature. In order to obtain a single-block-sparse estimate of the unknown echo path, a new least mean squares (LMS) algorithm is proposed by introducing the penalty of single block sparsity, which is the difference between the mixed $l_{2,1}$ norm and l_2 norm of the uniformly partitioned filter tap-weight vector, into the original mean-square-error cost function. This is motivated by the fact that the difference between the mixed $l_{2,1}$ norm and l_2 norm of a vector is minimised only when there is at most one non-zero block in the vector. Numerical simulation results show that the proposed algorithm can effectively estimate and track the unknown echo path, outperforming existing block-sparsity-induced LMS algorithms.

Introduction: In the application of network echo cancellation, a network echo canceller is used to model the network echo path, so that the echo signal can be removed from the near-end signal. Network echo path impulse response is sparse and single-clustered in nature. Network echo cancellers are required to have lengths of 512 or 1024, in order to identify a low-level echo [1]. However, the maximum dispersion time measured from North American telephone calls is 12 ms, which corresponds to 96 samples at 8 kHz sampling rate. Moreover, double reflections are seldom observed, hardly are more reflections than two measured [2].

Proportionate-type [3] and zero-attracting-type [4, 5] block-sparsity-induced least mean squares (LMS) algorithms have been proposed in the literature to identify block-sparse impulse responses, which are aimed at accelerating the convergence process of large and small coefficients groups, respectively. In view of the fact that the majority of groups in block-sparse systems are zero coefficients groups, the zero-attracting-type algorithms achieve faster overall convergence rates. The basis of the zero-attracting-type algorithms is to insert a penalty of block sparsity, such as a mixed $l_{2,1}$ norm [4], or an approximate mixed $l_{2,0}$ norm [5] of the uniformly divided filter tap-weight vector, into the cost function of the standard LMS algorithm. However, both of these penalties fail to make use of the characteristics of single block sparsity. Motivated by the fact that the difference between the mixed $l_{2,1}$ norm and l_2 norm of a vector is minimised only when there is at most one non-zero block in the vector, a new penalty of single block sparsity is proposed in this Letter. Then, a new block-sparsity-aware LMS algorithm is developed for network echo cancellers based on the proposed single-block-sparse penalty. Simulation results show that the proposed algorithm can effectively estimate and track the unknown echo path, outperforming existing block-sparsity-aware LMS algorithms.

Problem statement: In the context of network echo cancellation, the echo path, $\mathbf{w}^o = [w_1^o, w_2^o, \dots, w_L^o]^T$ with L being the echo path length, is excited by the far-end signal $\mathbf{u}(i) = [u_1(i), u_2(i), \dots, u_L(i)]$, where the notation (i) denotes the time index. The near-end signal $d(i)$ is the sum of the echo and the local noise $v(i)$:

$$d(i) = \mathbf{u}(i)\mathbf{w}^o + v(i) \quad (1)$$

where all the random variables $\{d(i), \mathbf{u}(i), v(i)\}$ are assumed to have zero means, and satisfy the following stationary model assumptions [6, p. 341]: (i) the processes $\{d(i), \mathbf{u}(i)\}$ are jointly wide-sense stationary; (ii) the noise sequence $\{v(i)\}$ is white and independent of $\mathbf{u}(j)$ for all i, j ; (iii) the auto-covariance matrix $\mathbf{R}_u = \mathbb{E}\mathbf{u}(i)\mathbf{u}(i)^T$ is positive definite.

The echo canceller $\mathbf{w}(i)$ is fed with the same far-end signal, and is used to estimate the echo path $\mathbf{w}^o$, so that the return signal $e(i)$ from near end to far end

$$e(i) = d(i) - \mathbf{u}(i)\mathbf{w}(i-1) \quad (2)$$

is close enough in the least-mean-square-error sense to the local noise $v(i)$.

Network echo path is a typical single-block-sparse impulse response, which consists of a pure delay, a significant part and a tail [7]. The significant part is a cluster of significantly large coefficients, while the other parts consist of coefficients that are just zero or of unnoticeably small values [1]. Then, in order to best estimate a network echo path, we need to find an estimator that is of a single-block-sparse structure.

Proposed algorithm: We firstly divide the echo canceller $\mathbf{w}(i)$ uniformly into N groups, such that there are $M = L/N$ coefficients in

each group. Let $\mathbf{w}_{[n]}(i) = [w_{(n-1)M+1}(i), w_{(n-1)M+2}(i), \dots, w_{nM}(i)]$ be the sub-vector for the n th group. Then, the mixed $l_{2,1}$ norm of $\mathbf{w}(i)$, $\|\mathbf{w}(i)\|_{2,1}$, is defined as

$$\|\mathbf{w}(i)\|_{2,1} = \sum_{n=1}^N \|\mathbf{w}_{[n]}(i)\|_2 \quad (3)$$

where $\|\cdot\|_2$ denotes the l_2 norm of a vector. Similarly, the l_2 norm of the echo canceller $\mathbf{w}(i)$, $\|\mathbf{w}(i)\|_2$, can also be defined in terms of these sub-vectors $\{\mathbf{w}_{[n]}(i)\}$

$$\|\mathbf{w}(i)\|_2 = \left(\sum_{n=1}^N \|\mathbf{w}_{[n]}(i)\|_2^2 \right)^{1/2} \quad (4)$$

We now proceed to consider the relationship between $\|\mathbf{w}(i)\|_{2,1}$ and $\|\mathbf{w}(i)\|_2$. Take the case when the vector $\mathbf{w}(i)$ is divided into $N = 3$ groups as an example. Assume that we have a cuboid whose length, width, and height are $\|\mathbf{w}_{[1]}(i)\|_2$, $\|\mathbf{w}_{[2]}(i)\|_2$, and $\|\mathbf{w}_{[3]}(i)\|_2$, respectively, with some measurement unit. Then, as shown in Fig. 1, $\|\mathbf{w}(i)\|_{2,1}$ will be the sum of the length, width, and height, while the $\|\mathbf{w}(i)\|_2$ will be the length of the space diagonal. Thus, it holds that

$$\|\mathbf{w}(i)\|_{2,1} - \|\mathbf{w}(i)\|_2 \geq 0 \quad (5)$$

where the equality holds if, and only if, there is at most one non-zero value among $\|\mathbf{w}_{[1]}(i)\|_2$, $\|\mathbf{w}_{[2]}(i)\|_2$, and $\|\mathbf{w}_{[3]}(i)\|_2$. That is to say, the difference $\|\mathbf{w}(i)\|_{2,1} - \|\mathbf{w}(i)\|_2$ can be minimised if, and only if, there is at most one non-zero block in the vector $\mathbf{w}(i)$. In view of this fact, we thus proceed to use the term $\|\mathbf{w}(i)\|_{2,1} - \|\mathbf{w}(i)\|_2$ as a penalty of single block sparsity.

Fig. 1 Comparison between $\|\mathbf{w}(i)\|_{2,1}$ and $\|\mathbf{w}(i)\|_2$ for any vector $\mathbf{w}(i)$ which is divided into three groups

Now, we need to seek the optimal solution that minimises the following cost function:

$$J(\mathbf{w}) = \frac{1}{2} \mathbb{E}[(d(i) - \mathbf{u}(i)\mathbf{w})^2] + \gamma(\|\mathbf{w}\|_{2,1} - \|\mathbf{w}\|_2) \quad (6)$$

where the penalty of single block sparsity is integrated with the traditional mean-square-error (MSE) cost function in order to obtain a single-block-sparse solution. The parameter $\gamma > 0$ balances the penalty term against the mean-square-error term.

Then, we start with the initial condition $\mathbf{w}(0) = 0$, and update the network echo canceller iteratively by following the stochastic-gradient-descent algorithm:

$$\begin{aligned} \mathbf{w}_l(i) &= \mathbf{w}_l(i-1) - \mu \widehat{\nabla}_{\mathbf{w}_l} J(\mathbf{w}(i-1)) \\ &= \mathbf{w}_l(i-1) + \mu u_l(i)e(i) - \frac{\kappa w_l(i-1)}{\|\mathbf{w}_{[\lceil l/M \rceil]}(i-1)\|_2 + \delta} + \frac{\kappa w_l(i-1)}{\|\mathbf{w}(i-1)\|_2 + \delta} \end{aligned} \quad (7)$$

for any $1 \leq l \leq L$, and where the notation $\widehat{\nabla}_{\mathbf{w}_l}(\cdot)$ denotes the stochastic gradient of $J(\mathbf{w})$ with respect to the l th entry of $\mathbf{w}$, w_l . Moreover, the parameter $\mu > 0$ is the step size, the error signal $e(i)$ is defined in (2), $\kappa = \mu\gamma$, the notation $\lceil \cdot \rceil$ denotes the ceiling function, and the parameter $\delta > 0$ avoids a division by zero.

It is clear from (7) that the proposed algorithm introduces a block-zero-attraction term, i.e.

$$- \kappa w_l(i-1) \left(\frac{1}{\|\mathbf{w}_{[\lceil l/M \rceil]}(i-1)\|_2 + \delta} - \frac{1}{\|\mathbf{w}(i-1)\|_2 + \delta} \right)$$

into the update equation of the original LMS algorithm. Now, we proceed to give some insights into the block-zero-attraction term.

Firstly, the zero-attraction term is zero in the case when $w_l(i-1) = 0$ or $\|w_{[l/M]}(i-1)\|_2 = \|w(i-1)\|_2$, which holds when $w(i-1) = 0$ or $w_{[l/M]}(i-1)$ is the only non-zero block in $w(i-1)$. In the other cases, the zero-attraction term attracts $w_l(i-1)$ to zero, in view of the fact that $\|w_{[l/M]}(i-1)\|_2 < \|w(i-1)\|_2$. Moreover, for coefficients in the same group, stronger zero-attraction forces are exerted on coefficients with larger magnitudes. That is because the difference in parentheses is a constant in this case. Then, for coefficients in large and small coefficients groups, we conclude, as shown in the following arguments, that generally weaker attractive forces will be exerted on coefficients in large coefficients groups than on those in small coefficients groups. We firstly consider the coefficients that have the same magnitude but are divided into different groups. Then, the coefficients inside large coefficients groups will receive weaker attractive forces, since values of the first term in parentheses are smaller for large coefficients groups. Now, we rewrite the zero-attraction term in the following form:

$$-\kappa \text{sgn}(w_l(i-1)) \left(\frac{|w_l(i-1)|}{\|w_{[l/M]}(i-1)\|_2 + \delta} - \frac{|w_l(i-1)|}{\|w(i-1)\|_2 + \delta} \right)$$

where the notation $\text{sgn}(\cdot)$ denotes the sign function. Then, for coefficients in different groups but sharing the same ratio in the first term in parentheses, weaker attractive forces will be exerted on larger coefficients, since values of the second term in parentheses will be larger for those coefficients.

Fig. 2 Unknown echo path (top: initial echo path; bottom: echo path after 27,000 iterations)

Fig. 3 MSD learning curves of the LMS, GZA-LMS, BS-LMS, BS-PNLMS, BS-IPNLMS, and proposed algorithm

Simulation results: In the simulation, we tested the tracking performance of the proposed algorithm. For comparison, performance of the LMS, block-sparse LMS (BS-LMS) [5], group-zero-attracting LMS (GZA-LMS) [4], block-sparse proportionate normalised LMS (BS-PNLMS), and its improved version BS-IPNLMS [3] were tested as well. Performance of these algorithms were evaluated in terms of mean-square deviation (MSD), which is defined as

$$\text{MSD}(i) = \mathbb{E} \|w^o - w(i)\|^2 \quad (8)$$

The simulation settings were as follows: entries of the excitation signal were drawn from the standard normal distribution. The unknown echo path was modelled by model 2 of international telecommunication union telecommunication standardisation sector recommendation G.168 [2] with a full length of $L = 1024$. As shown in Fig. 2, the echo path was initially generated with an echo return loss of 6.55 dB, and a delay

of 256 taps. Then, at the 27,000th iteration, the delay was increased to 640 taps, with magnitudes of non-zero coefficients decreased by 18.45 dB. The noise signal, uncorrelated with the excitation signal, was generated from a Gaussian distribution with mean 0 and variance 1.00×10^{-5} , which made a signal-to-noise ratio of 42 dB initially, and then 24 dB after 27,000 iterations. In the LMS algorithm, the step-size was set to $\mu = 0.42/L$. The group size was set to $M = 64$ in all the tested block-sparsity-aware LMS algorithms. In both proportionate-type algorithms, the step-size $\mu = 0.42$, and the parameter $\delta = 0.01/L$, which avoids a division by zero. Moreover, in the BS-PNLMS algorithm, we set $q = 0.01$, which prevents the initial adaptation of filter coefficients from stalling, and $\rho = 0.08$, which is utilised to set the step-sizes for small coefficients groups. In the BS-IPNLMS algorithm, the combination parameter was set to $\alpha = 0$. In the GZA-LMS algorithm, we set $\mu = 0.5$ and $\kappa = 5.43 \times 10^{-8}$. In both the BS-LMS and the proposed algorithm, the step size was set to $\mu = 0.8/L$, and the parameter $\delta = 10^{-8}$ to avoid a division by zero. In addition, in the BS-LMS algorithm, we set the parameter $\kappa = 8.00 \times 10^{-8}$, and $\alpha = 3$ which determines the attraction range. In the proposed algorithm, the parameter κ was set to 1.01×10^{-6} . These parameters were set so that all the tested algorithms converged to the same MSD level initially at their fastest possible speed. Fig. 3 shows the MSD learning curves of all these algorithms, where the results were averaged over 100 independent runs. It is clear from that figure that the proposed algorithm shows the best tracking performance in this case, which converges at the fastest speed to the same MSD level as the other algorithms in the initial stage, and then converges to a lower steady-state MSD value at a comparable speed after the echo path varies.

Conclusion: A new measure of single block sparsity was proposed, which is the difference between the mixed $l_{2,1}$ norm and l_2 norm of a vector. Then, a new block-sparsity-aware LMS algorithm was developed for the application of network echo cancellation by introducing the penalty of single block sparsity into the original MSE cost function. We tested and compared performance of the proposed algorithm and other existing block-sparsity-induced LMS algorithms in the literature. Simulation results show that the proposed algorithm can best track and estimate the unknown network echo path.

Acknowledgments: The work of Y. Wei was supported in part by a Chinese government scholarship. The work of Y. Zhang was supported in part by the National Natural Science Foundation of China (grant no. 61773133) and the Natural Science Foundation of Heilongjiang Province (grant no. F2016008).

© The Institution of Engineering and Technology 2018

Submitted: 29 March 2018 E-first: 15 June 2018

doi: 10.1049/el.2018.1065

One or more of the Figures in this Letter are available in colour online.

Ye Wei and Yonggang Zhang (College of Automation, Harbin Engineering University, Harbin, Heilongjiang 150001, People's Republic of China)

✉ E-mail: zhangyg@hrbeu.edu.cn

Chengcheng Wang (School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore)

References

- Deng, H., and Doroslovacki, M.: 'Proportionate adaptive algorithms for network echo cancellation', *Trans. Signal Process.*, 2006, **54**, (5), pp. 1794–1803
- 'Digital network echo cancellers', *International Telecommunication Union (ITU-T) Recommendation G.168*, 2012
- Liu, J., and Grant, S.L.: 'Proportionate adaptive filtering for block-sparse system identification', *IEEE/ACM Trans. Audio Speech Lang. Process.*, 2016, **24**, (4), pp. 623–630
- Chen, Y., Gu, Y., and Hero, A.O.: 'Regularized least-mean-square algorithms', *arXiv preprint arXiv:1012.5066*, 2010
- Jiang, S., and Gu, Y.: 'Block-sparsity-induced adaptive filter for multi-clustering system identification', *Trans. Signal Process.*, 2015, **63**, (20), pp. 5318–5330
- Sayed, A.H.: 'Adaptation, learning, and optimization over networks', *Found. Trends Mach. Learn.*, 2014, **7**, (4–5), pp. 311–801
- Gu, Y., Chen, Y., and Tang, K.: 'Network echo canceller with active taps stochastic localization', *Proc. IEEE ISIT*, 2005, **1**, pp. 556–559