Entropy-based stochastic gradient algorithm with adaptive neuron slope for all-pole whitening

V.R. Krstić[✉]

The stochastic gradient algorithm presented employs the time-varying neuron's slope to optimise performing of the all-pole filter-whitener maximising the joint Shannon's entropy. By using the time adaptive slope, which matches the unknown probability density function of an input process, a neuron slope selection issue of the original algorithm is facilitated and its tracking of non-stationary statistics is improved. The performing of algorithm is verified using the whitener as a front-end amplitude equaliser of the blind equaliser.

Introduction: Among the several algorithms suggested for parameter estimation in a linear all-poles model, the extended least mean square (ELMS) algorithm is probably the most prominent one because it enjoys a low computation cost and self-stabilisation property [1, 2]. Its simplicity comes from the fact that it has no gradient term based on a cost function optimisation but it constructs the update term using the currently available output signal. Unfortunately, the lack of a real gradient makes the performance of the ELMS algorithm inferior to the most faded spectral components of a signal applied at the predictor input.

The algorithm presented in this Letter stems from the stochastic gradient algorithm maximising the joint Shannon's entropy of equaliser outputs [3] according to blind deconvolution approach by Bell and Sejnowski [4]. The specific property of the original algorithm is the user-definable parameter which matches the slope of the activation (neuron) function to the unknown probability density function (PDF) of an input signal. Aiming to facilitate a slope selection issue and to achieve a better tracking of non-stationary statistics, we have improved the updating term of the algorithm by introducing the a posteriori adaptive neuron slope in combination with the coefficient-leaky term.

Joint entropy maximisation (JEM) whitener model: This section presents the basic model of the decision feedback equaliser (DFE) optimising the JEM cost and its reduction to the corresponding pure recursive linear predictor (whitener). The basic model of JEM-DFE in Fig. 1 can be described as follows: data symbols $\{a_n\}$ applied to the input of a noiseless channel are a zero-mean sequence of independent identically distributed variables with a finite variance and sub-Gaussian distribution. The DFE output $z_n = x_n - \boldsymbol{b}_n^T \boldsymbol{r}_n$ at the time $t = nT$, T is a symbol period, is the sum of channel output x_n and the convolution sum of vectors $\boldsymbol{b}_n = [b_{n,1}, \ \ldots, \ b_{n,N}]^{\mathrm{T}}$ and $\boldsymbol{r}_n = [r_{n-1}, \ \ldots, r_{n-N}]^{\mathrm{T}}$ representing, respectively, DFE filter coefficients $\{b_j\}$ and neuron outputs $r_n = g(z_n)$. In this DFE model, the neuron $g(\cdot)$ is a monotone (saturating) function whose slope g' determines the DFE efficiency of information extraction.

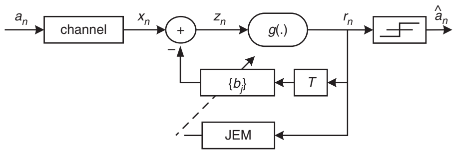

Fig. 1 *Basic model of JEM decision feedback equaliser*

The basic JEM stochastic gradient algorithm given by Krstić and Dukić [3]

$$b_{n+1,j} = b_{n,j} - \mu z_n \big(1 - \beta |z_n|^2\big) r_{n-j}^*, \quad j = 1, \ \ldots, \ N \tag{1}$$

has been derived for the parametric complex function $g(z_n, \ \beta) = z_n(1 + \beta |z_n|^2)$ of complex-valued variable z_n assuming the previously detected symbols are correct, i.e. $r_{n-j} = a_{n-j}$, where μ is a step size, β is a real positive parameter (slope) and * denotes complex conjugation. The corresponding JEM-type whitening (JEM-W) algorithm, which is our focus in this Letter, is a result of heuristic linearisation of the equaliser achieved by dropping the nonlinear mapping of its outputs, i.e. $g(z_n \beta) = z_n$ and given by $z_n = x_n - \boldsymbol{b}_n^T \boldsymbol{z}_n$

$$b_{n+1,j} = b_{n,j} - \mu_W z_n \big(1 - \beta_W |z_n|^2\big) z_{n-j}^*, \quad j = 1, \ \ldots, \ N \tag{2}$$

where $z_n = [z_{n-1}, \ \ldots, \ z_{n-N}]^{\mathrm{T}}$ and μ_W is a step size. It should be noted that the algorithm in (2) has retained the entropic character of the error

$$e_n = z_n(1 - \beta_W |z_n|^2) \tag{3}$$

which can be seen as an a priori prediction error, and in the specific case for $\beta_W = 2/|z_n|^2$ the JEM-W is reduced to the ELMS.

As it is mentioned in the 'Introduction' section, the selection of the slope is a key issue of JEM-DFE design. In contrast to the basic JEM algorithm in (1) where the slope β is inversely proportional to the variance of an input process [4] and can be seen as a statistical constant of the applied signal [3, 5], the selection of the slope β_W in (2) is a tricky problem because there is no saturation in whitener outputs. Namely, to match a PDF with a larger variance corresponding to channel outputs with deep spectral nulls, the slope's decrease results in reduced influence of the nonlinear term $\beta|z_n|^2$ in (2) so that the JEM-W algorithm loses ability to extract the most faded components of a power spectrum. On the other hand, when the slope increases the JEM-W algorithm demonstrates a fast convergence rate and good spectral estimation but it can be followed by some whitener poles located inside but very close to the unit circle in the z-plane. This difficulty has been our motivation to introduce an adaptive slope (ASL) into the JEM-W algorithm with the following adjustment scenario: at the beginning, the slope would be large enough to provide fast and stable estimation and when the whitener approaches a steady state the slope would be small enough to prevent its instability.

JEM-W algorithm with ASL: The JEM-W algorithm with the ASL (JEM-ASL)

$$b_{n+1,j} = (1 - \gamma) b_{n,j} - \mu_W z_n \big(1 - \beta_n |z_n|^2\big) z_{n-j}^*, \quad j = 1, \ \ldots, \ N \tag{4}$$

introduces the leaky term γb_n, where γ is a leaky factor, and a time-variable slope β_n (subscript w is dropped for simplicity). As is well known, the leaky term gradually decreases the coefficient norm $\|\boldsymbol{b}_n\|$, and it is commonly used in linear regression models with different motivations: to regularise their transient behaviour, stability in a finite precision implementation is improved and different undesirable effects (see references in [6]) are reduced. On the other hand, the estimates achieved in the presence of leaky are biased so that some trade-off is required. The leaky in (4) is introduced for two reasons: first, acting in opposition to the entropic term $\beta|z_n|^2$ it prevents unconstrained growth of predictor coefficients [5]. In fact, it enables using larger values of the slope without the risk of instability. Secondly, its efficiency to regularise the transient behaviour of the algorithm we have used as a measure to decide whether to increase or decrease the slope β_n.

The adaptation of the slope β_n is a posteriori based. It means that the predictor, at the time $t = nT$ but before the next updated input x_{n+1}, calculates the coefficient vector $\boldsymbol{b}_{n+1}^{\mathrm{ASL}}$ with $(\gamma > 0)$ and $\boldsymbol{b}_{n+1}^{\mathrm{AS}}$ without $(\gamma = 0)$ leaky and the corresponding a posteriori outputs $(\tilde{u}_n^{\mathrm{ASL}}, \ \tilde{u}_n^{\mathrm{AS}})$ and errors $(\tilde{e}_n^{\mathrm{ASL}}, \ \tilde{e}_n^{\mathrm{AS}})$. Furthermore, based on the difference between modules of a posteriori errors the predictor decides whether to increase or decrease the slope β_n according to the rule: if $|\tilde{e}_n^{\mathrm{ASL}}| \le |\tilde{e}_n^{\mathrm{AS}}|$ the slope decreases and if $|\tilde{e}_n^{\mathrm{ASL}}| > |\tilde{e}_n^{\mathrm{AS}}|$ the slope increases.

The adaptation of the slope is implemented in a similar manner as it was done in [6] by the following if–else relation:

$$\begin{aligned} &\text{if } |\tilde{e}_n^{\mathrm{ASL}}| > |\tilde{e}_n^{\mathrm{AS}}| \text{ then} \\ &\text{set } m_{n+1} = \max(m_n - l_d, \ 0) \\ &\text{else} \\ &\text{set } m_{n+1} = \min(m_n + l_u, \ M) \\ &\text{end if} \end{aligned} \tag{5}$$

and using the new quantised function

$$\beta = f(m) = \alpha_0 - \alpha_1(m_n/M) \tag{6}$$

which calculates the slope β_{n+1} given the current slope β_n, where $m_n = 0, \ \ldots, \ M$ is an independent variable, and $(M, \ l_d, \ l_u) \in \mathbb{Z}$, $(\alpha_0, \ \alpha_1) \in \mathbb{R}$, m_0 are user-definable parameters.

Performance of JEM-ASL algorithm: To verify JEM-ASL efficiency, we have employed the DFE equaliser which in the blind mode acts as a linear cascaded $T/2$ fractionally spaced equaliser [FSE-constant modulus algorithm (CMA)] including the gain control device (GC), the JEM whitener (WT) and the FSE equaliser (TE) controlled by the

CMA [3, 5]. Fig. 2 presents the FSE-CMA equaliser whose filters WT and TE have lengths $N = 5$ and $L = 24$, respectively, and the initial values of their coefficients are all zero except of the TE central coefficients $c_{1,r} = c_{2,r} = 1.0$. The user-definable parameters in (5) and (6) are $\{\alpha_0 = 1.5, \alpha_1 = 1.3, M = 1000, l_d = 5, l_u = 15, m_0 = 1\}$ and the leaky factor is $\gamma = 2^{-13}$. The given parameters are selected in such a way to ensure a fast variation of the slope in the wide range from 0.2 to 1.5. The performing of JEM-ASL is given in comparison with JEM-W for slopes $\beta_W = (0.2, 1.5)$, and simulations are carried out for (16, 64, 128)-QAM signals transmitted over the severe multipath channel Mp-E with signal-to-noise ratio of 30 dB; the magnitude response of the multipath (Mp) channel is available in [5]. The results presented in Fig. 3 are given in terms of convergence of the whitener vector norm $\|\boldsymbol{b}_n\|$ and the mean square error (MSE) of the FSE-CMA output y_n. As can be seen in Fig. 3*a*, the JEM-ASL whitener presets a fast and stable initial convergence and a strong saturation in the steady-state regime independently of the applied signal. Consequently, such whitener behaviour speeds up the convergence of TE, Fig. 3*b*. On the other hand, the convergence of the JEM-W whitener and, hence, TE equaliser significantly depends on the selected slope β_W. Besides, the unconstrained growth of the coefficient norm of the JEM-W whitener can be risky for its stability.

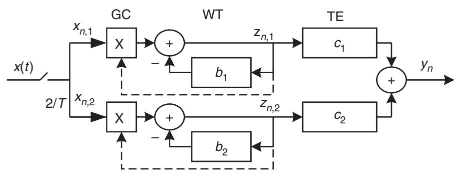

Fig. 2 *FSE-CMA equaliser with front-end JEM whitener*

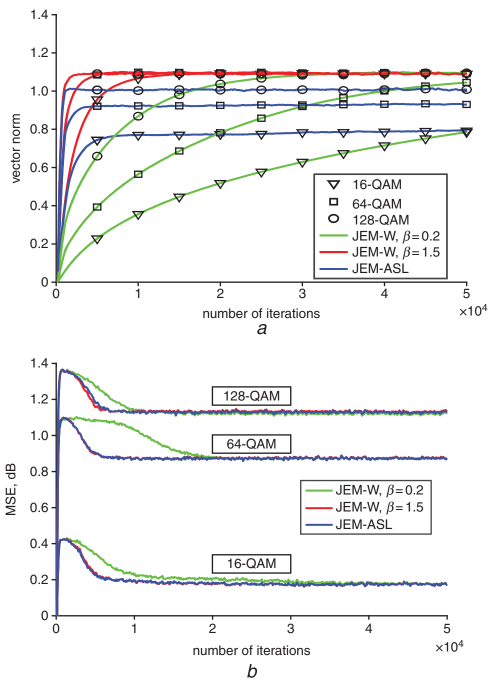

Fig. 3 *Behavior of the JEM whitener and its influence on the FSA-CMA convergence for (16, 64, 128)-QAM signals*

a Convergence characteristics of whitener vector norm
b MSE of FSE-CMA averaged over 100 runs 16-QAM ($\mu_W = 2^{-18}$), 64-QAM ($\mu_W = 2^{-22}$) and 128-QAM ($\mu_W = 2^{-23}$)

Conclusion: Using the variable neuron slope, the JEM-ASL algorithm eliminates a parameter selection issue and in such a way facilitates design of the whitener and improves its performance. According to the simulation results, the JEM-AS algorithm provides fast and stable convergence of the whitener and, hence, equaliser independently of applied QAM signals.

Submitted: *28 August 2015*
doi: 10.1049/el.2015.3052
One or more of the Figures in this Letter are available in colour online.

V.R. Krstić (*Institute 'Mihajlo Pupin', Communications Department, Volgina 15, 11060 Belgrade, Serbia*)

✉ E-mail: vladimir.krstic@pupin.rs

References

1 Macchi, O.: 'Adaptive processing: the least mean squares approach with applications in transmission' (Wiley & Sons, New York, 1995)
2 Lopez-Valcarce, R.: 'Analysis of a first-order adaptive recursive predictor'. Proc. 44th IEEE Conf. on Decision and Control, and the European Control Conf., Spain, 12–15 December 2005, pp. 4791–4796
3 Krstić, V.R., and Dukić, M.L.: 'Blind DFE with maximum-entropy feedback', *IEEE Signal Process. Lett.*, 2009, **16**, (1), pp. 26–29
4 Bell, A.J., and Sejnowski, T.J.: 'An information-maximization approach to blind separation and blind deconvolution', *Neural Comput.*, 1995, **7**, pp. 1129–1159
5 Krstić, V.R., and Dukić, M.L.: 'Decision feedback blind equalizer with tap-leaky whitening for stable structure-criterion switching', *Int. J. Digit. Multimedia Broadcast.*, 2014, **2014**, (2014), p. 10, Article ID 987039. Available at http://www.dx.doi.org/10.1155/2014/987039
6 Kamenetsky, M., and Widrow, B.: 'A variable leaky LMS adaptive algorithm'. Proc. 38th Asilomar Conf. on Signal, Systems and Computers, Pacific Grove, California, November 2004, vol. 1, pp. 125–128