Adaptive Digital Equalization in Optical Coherent Receivers With Stokes-Space Update Algorithm

Adaptive Digital Equalization in Optical Coherent Receivers with Stokes-Space Update Algorithm

Monica Visintin, *Member, IEEE*, Gabriella Bosco, *Senior Member, IEEE*, Pierluigi Poggiolini, *Member, IEEE*, Fabrizio Forghieri, *Member, IEEE*

***Abstract*—In this paper we describe a novel update algorithm for the filter coefficients of the adaptive digital equalizer in coherent receivers, which is based on error signals evaluated in Stokes space and is insensitive to both phase-noise and frequency-offset. We also introduce an optimized decision rule in the Stokes space, which takes into account the exact statistics of noise and yields a performance improvement with respect to the minimum distance decision criterion. We compare the performance of the new algorithm to the standard constant-modulus algorithm (CMA) for polarization-multiplexed (PM) 16QAM modulation, achieving similar performance in the absence of phase noise, with comparable complexity. Differently from CMA, the proposed Stokes-space algorithm (SSA) allows to remove the phase offset between polarizations, thus enabling the use of a joint carrier-phase estimation algorithm on both polarizations, which in turns yields a nearly doubled phase noise tolerance.**

***Index Terms*—Coherent communications, Modulation, Fiber optics links and subsystems.**

I. Introduction

In optical systems based on coherent detection, linear system impairments can be efficiently compensated for in the digital domain through digital signal processing (DSP) algorithms. Typically, a scalar FIR filter removes the bulk of the fiber chromatic dispersion (CD) accumulated through propagation. Then an adaptive butterfly equalizer recovers the polarization of the transmitted signal and compensates for polarization effects (such as polarization dependent loss and polarization mode dispersion), bandwidth limitations and residual CD. Several methods have been proposed to update the adaptive equalizer taps, the most popular being based on the least-mean-square (LMS) and constant-modulus algorithms (CMA) [1].

The advent of coherent detection has enabled the use of multi-level modulation formats in optical transmission [2], significantly increasing spectral efficiency. However, high-order modulation formats are more sensitive to laser phase noise effects, thus particular attention has to be paid to the design of proper carrier phase estimation (CPE) algorithms. An important advantage of CMA over LMS stems from the fact that it is insensitive to phase noise: this fact allows the butterfly equalizer and the CPE block to operate independently, avoiding the need of the tricky and complex feed-back loops which are instead required when using the LMS update algorithms.

Recently, the potential advantages given by processing the received signal samples using a representation in the Stokes space have been highlightd by several groups [3]-[7]. In [8], [9] we proposed and demonstrated an alternative method for the updates of the butterfly equalizer taps, based on an error signal evaluated in Stokes space. This method shares with CMA the advantage of being insensitive to phase noise. In addition, differently from the other techniques, it automatically guarantees that the constellations on the two polarizations are aligned in phase, thus allowing to apply carrier phase noise compensation by averaging the phase estimation over the two polarizations, yielding a higher robustness to phase noise. Preliminary results were reported in [8], [9]. In this paper, we describe in detail the technique, comparing it to the CMA algorithm in terms of both performance and complexity. Section II reports the system model and notation that will be used throughout the paper. In Section III the Stokes space algorithm (SSA) is detailed, whilst, to enable a complexity comparison, Section IV reports the equations used in the update rule when either the CMA or LMS criterion is used. Section V is devoted to the description of the decision rule used in the Stokes space. Finally, in Section VI simulation results are shown, comparing the performance of receivers based on SSA and CMA when a polarization multiplexed (PM) 16-points quadrature-amplitude modulation (16QAM) is used, both in terms of back-to-back sensitivity and tolerance to phase noise. Robustness to PDL is also investigated. Conclusions on the pros and cons of the proposed system are finally drawn in Section VII.

II. System model

The transmitter (Tx) generates a polarization-multiplexed (PM) signal, where the complex envelope of the $\hat{x}$ and $\hat{y}$ polarized electric fields can be written as:

$$E_x(t) = \sqrt{k \cdot P_T} \sum_{n=-\infty}^{\infty} s_{n,x} q(t - nT)$$
$$E_y(t) = \sqrt{k \cdot P_T} \sum_{n=-\infty}^{\infty} s_{n,y} q(t - nT) \qquad (1)$$

being P_T the total average power of the dual polarization signal, k a normalization constant which depends on the modulation format, $q(t)$ the transmitted pulse with unitary energy,

M. Visintin, G. Bosco, and P. Poggiolini are with the Electronic and Telecommunication Department, Politecnico di Torino, Italy (e-mail: gabriella.bosco@polito.it). F. Forghieri is with Cisco Photonics Italy srl (e-mail: fforghie@cisco.com). This work was supported by Cisco Systems within a SRA contract.

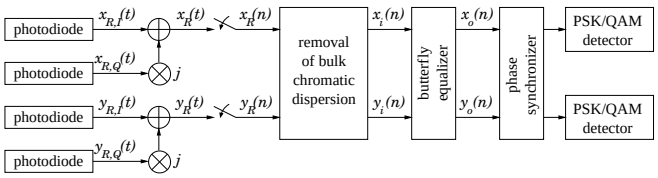

Fig. 1. Structure of the receiver.

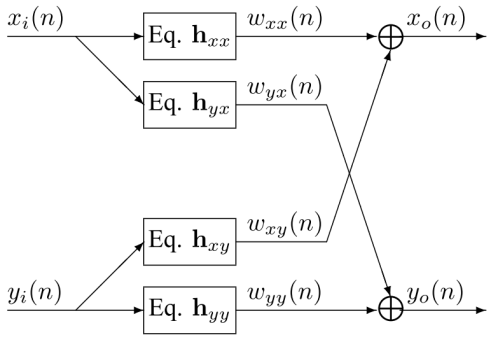

Fig. 2. Structure of the butterfly equalizer with inputs $x_i(n)$ and $y_i(n)$ and outputs $x_o(n)$ and $y_o(n)$.

T the symbol interval, $s_{k,x}$ and $s_{k,y}$ the complex constellation symbols (normalized such as the minimum distance is equal to 1). The symbol rate is $R_s = 1/T$. As an example, for QPSK, $k = 1/2$ and $s_{k,x}, s_{k,y}$ have real and imaginary parts belonging to the set $\{\pm 0.5\}$, whilst for 16QAM, $k = 2/5$ and $s_{k,x}, s_{k,y}$ have real and imaginary parts belonging to the set $\{\pm 1.5, \pm 0.5\}$.

The PM signal propagates over an optical transmission system, composed of several fiber spans with optical amplification, and is then detected by a standard coherent receiver (Rx) [10]. The signals (see Fig. 1) $x_{R,I}(t)$, $x_{R,Q}(t)$ $y_{R,I}(t)$ $y_{R,Q}(t)$ at the output of the photodiodes, which in an ideal case would correspond to the real and imaginary components of $E_x(t)$ and $E_y(t)$ in (1), respectively, are sampled by an analog-to-digital converter (ADC) at a sampling frequency $N_s \cdot R_s$, where N_s is the number of samples per symbol. Two sequences of complex numbers are obtained: $x_R(n)$ and $y_R(n)$, corresponding to the sampled versions the signals $x_R(t) = x_{R,I}(t) + jx_{R,Q}(t)$ and $y_R(t) = y_{R,I}(t) + jy_{R,Q}(t)$, respectively. The complex signal samples then go through a digital signal processing (DSP) chain [1], composed of: a bulk CD compensation stage; an adaptive butterfly equalizer; a carrier frequency and phase compensation stage.

The complex signal samples at the output of the butterfly equalizer (see Fig.2) can be written as:

$$\begin{aligned} x_o(n) &= \mathbf{x}_i^{(n)} \cdot \mathbf{h}_{xx}^{(n)} + \mathbf{y}_i^{(n)} \cdot \mathbf{h}_{xy}^{(n)} \\ y_o(n) &= \mathbf{x}_i^{(n)} \cdot \mathbf{h}_{yx}^{(n)} + \mathbf{y}_i^{(n)} \cdot \mathbf{h}_{yy}^{(n)} \end{aligned} \tag{2}$$

where $\mathbf{a} \cdot \mathbf{b}$ indicates the scalar product between vectors $\mathbf{a}$ and $\mathbf{b}$, and:

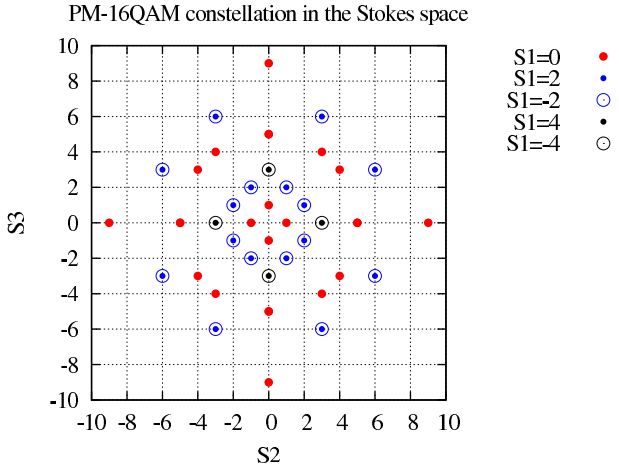

Fig. 3. Representation in the Stokes space of the ideal transmitted PM-16QAM constellation points for $P_T = 5$.

$$\begin{aligned} \mathbf{x}_i^{(n)} &= \left[x_i(n), x_i(n-1) \dots x_i(n-N+1)\right]^T \\ \mathbf{y}_i^{(n)} &= \left[y_i(n), y_i(n-1) \dots y_i(n-N+1)\right]^T \\ \mathbf{h}_{xx}^{(n)} &= \left[h_{xx}^{(n)}(1), h_{xx}^{(n)}(2) \dots h_{xx}^{(n)}(N)\right]^T \\ \mathbf{h}_{xy}^{(n)} &= \left[h_{xy}^{(n)}(1), h_{xy}^{(n)}(2) \dots h_{xy}^{(n)}(N)\right]^T \\ \mathbf{h}_{yx}^{(n)} &= \left[h_{yx}^{(n)}(1), h_{yx}^{(n)}(2) \dots h_{yx}^{(n)}(N)\right]^T \\ \mathbf{h}_{yy}^{(n)} &= \left[h_{yy}^{(n)}(1), h_{yy}^{(n)}(2) \dots h_{yy}^{(n)}(N)\right]^T \end{aligned} \tag{3}$$

where $\mathbf{a}^T$ is the transpose of vector $\mathbf{a}$. $x_i(n)$ and $y_i(n)$ are the complex signal samples at the input of the butterfly equalizer, whilst $h_{xx}^{(n)}(k), h_{yx}^{(n)}(k), h_{xy}^{(n)}(k)$ and $h_{yy}^{(n)}(k)$, with $k = 1, \dots, N$, are the impulse responses of 4 FIR filters of the butterfly equalizer at time n, having N taps each. We will assume in the following that the tap weights are updated once per symbol interval, specifically when n is an integer multiple of N_s.

III. The Stokes-space algorithm (SSA)

The signals $x_o(n)$ and $y_o(n)$ at the output of the butterfly equalizer can be represented in the Stokes space as:

$$\begin{cases} S_{1e}(n) = |x_o(n)|^2 - |y_o(n)|^2 \\ S_{2e}(n) = 2\Re\{x_o(n)y_o^*(n)\} = |x_o(n)|\,|x_o(n)|\cos(\phi_d) \\ S_{3e}(n) = 2\Im\{x_o(n)y_o^*(n)\} = |x_o(n)|\,|x_o(n)|\sin(\phi_d) \end{cases} \tag{4}$$

where $x_o(n) = |x_o(n)|e^{j\phi_x}$, $y_o(n) = |y_o(n)|e^{j\phi_y}$ and $\phi_d = \phi_x - \phi_y$. The couple of complex values $x_o(n)$ and $y_o(n)$ thus correspond to the real valued vector $\mathbf{S}_e(n) = [S_{1e}(n), S_{2e}(n), S_{2e}(n)]$. Note that the Stokes representation removes the information on the individual phase of each polarizations, leaving only the information on the differential phase ϕ_d. For this reason, different constellation points may be characterized by the same vector in Stokes space, thus the number of points in Stokes space is lower (approximately by a factor of 4 for QAM modulation) than

the number of constellation points. As an example, the 16 symbols of the PM-QPSK modulation with total power P_T are mapped into 4 points with Stokes representations $[0,1,0]P_T$, $[0,-1,0]P_T$,$[0,0,1]P_T$, $[0,0,-1]P_T$. The representation in Stokes space of PM-16QAM constellation is shown in Fig.3: the 16x16=256 points of the constellation are mapped into 60 points in the Stokes space.

Since the Stokes representation is independent of the absolute phase in the two polarizations, the adaptation of the equalizer tap weights based on a function of the Stokes parameters makes the equalizer insensitive to any form of phase noise or frequency offset, which will be removed by a cascaded, independently designed, block.

By substituting (2) into (4) and omitting, for sake of simplicity, the dependency on n, we get:

$$\begin{aligned}
S_{1e} &= |\mathbf{x}_i \cdot \mathbf{h}_{xx}|^2 + |\mathbf{y}_i \cdot \mathbf{h}_{xy}|^2 + 2\Re\{(\mathbf{x}_i \cdot \mathbf{h}_{xx})(\mathbf{y}_i \cdot \mathbf{h}_{xy})^*\} - \\
&\quad - \left[|\mathbf{y}_i \cdot \mathbf{h}_{yy}|^2 + |\mathbf{x}_i \cdot \mathbf{h}_{yx}|^2 + 2\Re\{(\mathbf{x}_i \cdot \mathbf{h}_{yx})(\mathbf{y}_i \cdot \mathbf{h}_{yy})^*\}\right] \\
S_{2e} &= 2\Re\{(\mathbf{x}_i \cdot \mathbf{h}_{xx})(\mathbf{x}_i \cdot \mathbf{h}_{yx})^*\} + 2\Re\{(\mathbf{x}_i \cdot \mathbf{h}_{xx})(\mathbf{y}_i \mathbf{h}_{yy})^*\} + \\
&\quad + 2\Re\{(\mathbf{y}_i \cdot \mathbf{h}_{xy})(\mathbf{x}_i \cdot \mathbf{h}_{yx})^*\} + 2\Re\{(\mathbf{y}_i \cdot \mathbf{h}_{xy})(\mathbf{y}_i \cdot \mathbf{h}_{yy})^*\} \\
S_{3e} &= 2\Im\{(\mathbf{x}_i \cdot \mathbf{h}_{xx})(\mathbf{x}_i \cdot \mathbf{h}_{yx})^*\} + 2\Im\{(\mathbf{x}_i \cdot \mathbf{h}_{xx})(\mathbf{y}_i \cdot \mathbf{h}_{yy})^*\} + \\
&\quad + 2\Im\{(\mathbf{y}_i \cdot \mathbf{h}_{xy})(\mathbf{x}_i \cdot \mathbf{h}_{yx})^*\} + 2\Im\{(\mathbf{y}_i \cdot \mathbf{h}_{xy})(\mathbf{y}_i \cdot \mathbf{h}_{yy})^*\}
\end{aligned} \tag{5}$$

The error function to be minimized by the equalizer is the difference between the equalized value $\mathbf{S}_e(n)$ and the transmitted value $\hat{\mathbf{S}}(n)$, which may be either known (in an initial training phase), or estimated through the decision rule is described in Sect V (when operating in decision-directed mode during the tracking phase):

$$\begin{aligned}
f(\mathbf{h}^{(n)}) &= f(\mathbf{h}_{xx}^{(n)}, \mathbf{h}_{xy}^{(n)}, \mathbf{h}_{yx}^{(n)}, \mathbf{h}_{yy}^{(n)}) = \\
&= \|\mathbf{S}_e(n) - \hat{\mathbf{S}}(n)\|^2 \\
&= (S_{1e}(n) - \hat{S}_1(n))^2 + (S_{2e}(n) - \hat{S}_2(n))^2 + \\
&\quad + (S_{3e}(n) - \hat{S}_3(n))^2
\end{aligned} \tag{6}$$

The equalizer taps are updated at each symbol interval (i.e. for n integer multiple of N_s), with the following updating rule, derived from the gradient algorithm [11]:

$$\begin{aligned}
\mathbf{h}_{xx}^{(n+1)} &= \mathbf{h}_{xx}^{(n)} - \mu \nabla_{\mathbf{h}_{xx}^{(n)}} f(\mathbf{h}^{(n)}) \\
\mathbf{h}_{xy}^{(n+1)} &= \mathbf{h}_{xy}^{(n)} - \mu \nabla_{\mathbf{h}_{xy}^{(n)}} f(\mathbf{h}^{(n)}) \\
\mathbf{h}_{yx}^{(n+1)} &= \mathbf{h}_{yx}^{(n)} - \mu \nabla_{\mathbf{h}_{yx}^{(n)}} f(\mathbf{h}^{(n)}) \\
\mathbf{h}_{yy}^{(n+1)} &= \mathbf{h}_{yy}^{(n)} - \mu \nabla_{\mathbf{h}_{yy}^{(n)}} f(\mathbf{h}^{(n)})
\end{aligned} \tag{7}$$

where μ is a small positive real number (updating coefficient). It can be shown (see Appendix VIII-A) that the gradients in Eq.(7) can be evaluated as:

$$\begin{aligned}
\nabla_{\mathbf{h}_{xx}^{(n)}} f(\mathbf{h}^{(n)}) &= C_1 \cdot \left(\mathbf{x}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yy}^{(n)}} f(\mathbf{h}^{(n)}) &= C_2 \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{xy}^{(n)}} f(\mathbf{h}^{(n)}) &= C_1 \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yx}^{(n)}} f(\mathbf{h}^{(n)}) &= C_2 \cdot \left(\mathbf{x}_i^{(n)}\right)^*
\end{aligned} \tag{8}$$

where a^* is the complex conjugate of a and:

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} a & b \\ b^* & -a \end{bmatrix} \begin{bmatrix} x_o(n) \\ y_o(n) \end{bmatrix} \tag{9}$$

being

$$\begin{aligned}
a &= 4[S_{1e}(n) - \hat{S}_1(n)] \\
b &= 4[(S_{2e}(n) - \hat{S}_2(n)) + j(S_{3e}(n) - \hat{S}_3(n))]
\end{aligned} \tag{10}$$

Initially, during the training phase, known data are transmitted, and the exact values of the Stokes representation of the transmitted symbols $\mathbf{S}(n)$ is used in place of $\hat{\mathbf{S}}(n)$ in the updating algorithm given in eqns. (7) and (8); during the tracking phase, the transmitted symbols are not known and decisions are made in the Stokes space to obtain $\hat{\mathbf{S}}(n)$, and the algorithm is applied as described above.

The overall block diagram of the adaptive equalizer, including the updating rule of Eq. (7), is reported in Fig. 4, while the generation of coefficients C_1 and C_2 of Eq. (9) is shown in Fig. 5.

IV. Comparison with CMA and LMS equalizer updating rules

In the case of CMA algorithm [1], two objective functions are iteratively minimized by the adaptive equalizer:

$$\begin{aligned}
f_{CMA,x}(\mathbf{h}(n)) &= (|x_o(n)|^2 - |\hat{x}(n)|^2)^2 \\
f_{CMA,y}(\mathbf{h}(n)) &= (|y_o(n)|^2 - |\hat{y}(n)|^2)^2
\end{aligned} \tag{11}$$

where $|\hat{x}(n)|$ and $|\hat{y}(n)|$ are the values of the modulus of the known transmitted symbols during the training phase, whereas in the tracking phase they are estimated by comparing $|x_o(n)$ and $|y_o(n)|$ with thresholds. Since $x_o(n)$ only depends on $\mathbf{h}_{xx}$ and $\mathbf{h}_{xy}$, and $y_o(n)$ only depends on $\mathbf{h}_{yy}$ and $\mathbf{h}_{yx}$, the updating rules for the filters is again that described in eqn. (7), but the gradients are

$$\begin{aligned}
\nabla_{\mathbf{h}_{xx}^{(n)}} f_{CMA,x}(\mathbf{h}^{(n)}) &= C_{1,CMA} \cdot \left(\mathbf{x}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yy}^{(n)}} f_{CMA,y}(\mathbf{h}^{(n)}) &= C_{2,CMA} \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{xy}^{(n)}} f_{CMA,x}(\mathbf{h}^{(n)}) &= C_{1,CMA} \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yx}^{(n)}} f_{CMA,y}(\mathbf{h}^{(n)}) &= C_{2,CMA} \cdot \left(\mathbf{x}_i^{(n)}\right)^*
\end{aligned} \tag{12}$$

which is similar to eq. (8), but the coefficients $C_{1,CMA}$ and $C_{2,CMA}$ are obtained as

$$\begin{bmatrix} C_{1,CMA} \\ C_{2,CMA} \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_o(n) \\ y_o(n) \end{bmatrix} \tag{13}$$

$$\begin{aligned}
a &= 2\left(|x_o(n)|^2 - |\hat{x}(n)|^2\right) \\
b &= 2\left(|y_o(n)|^2 - |\hat{y}(n)|^2\right)
\end{aligned} \tag{14}$$

Comparing Eqs. (13) and (9), it is evident that the complexity of the evaluation of the coefficients C_1 and C_2 is higher for the SSA than for CMA. However, most of the real-time computational complexity is due to the operations performed in the block diagram shown in Fig. 4, which is common to CMA and SSA, yielding to a similar overall complexity for

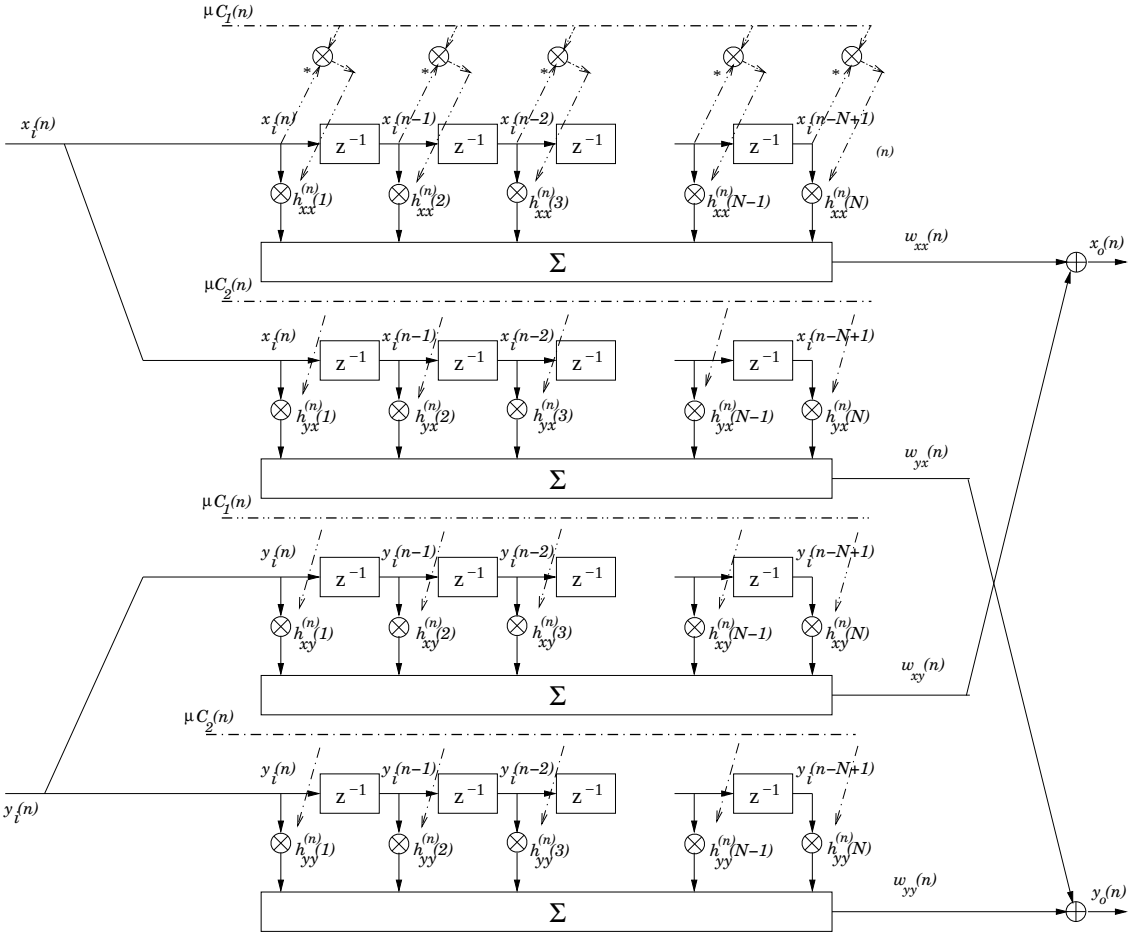

Fig. 4. Block diagram of the adaptive equalizer, including the updating rule: for example, $h_{xx}^{(n+1)}(3) = h_{xx}^{(n)}(3) - \mu C_1 x_i^*(n-2)$. The generation of coefficients C_1 and C_2 is shown in Fig. 5.

the two algorithms. We do not report here a more detailed complexity analysis, leaving it to further investigations, since it will be strongly dependent on the actual implementation.

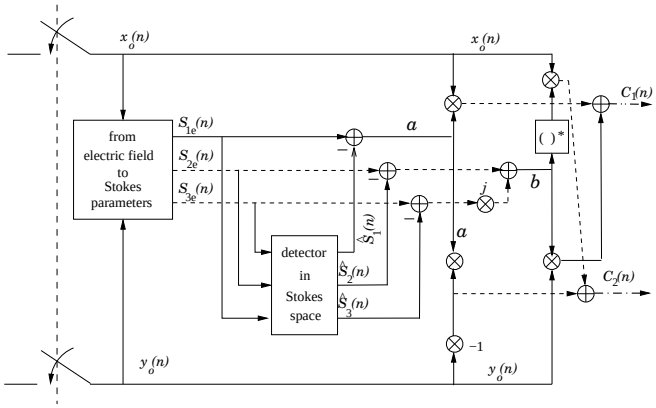

Fig. 5. Generation of coefficients C_1 and C_2 for the Stokes equalizer.

The CMA algorithm has the advantage of being insensitive to phase shifts (due to phase noise or frequency offset), but suffers from degeneracy problems since it is possible that during the tracking phase $\mathbf{h}_{xx}$ becomes equal to $\mathbf{h}_{yx}$ and $\mathbf{h}_{yy}$ becomes equal to $\mathbf{h}_{xy}$, so that $x_o(n) = y_o(n)$, and only one of the two polarizations is output by the equalizer. Also, the outputs of the CMA equalizers can be in general written as $x_o(n) = x(n - L_x)e^{j\theta_x(n)}$ and $y_o(n) = y(n - L_y)e^{j\theta_y(n)}$, with L_x and L_y integer positive numbers accounting for delays, which could be in general different. If $L_x \neq L_y$, a further block is needed to temporally align $x_o(n)$ and $y_o(n)$. Moreover, the two equalizers working separately on the two polarizations could converge to two constellations having a different phase rotation with respect to the original ones, so that two separate phase synchronizers are needed to remove the phase offsets $\theta_x(n)$ and $\theta_y(n)$.

In the case of LMS (Least Mean Squares) algorithm [1] the objective function to be minimized is

$$f_{LMS}(\mathbf{h}^{(n)}) = |x_o(n) - \hat{x}(n)|^2 + |y_o(n) - \hat{y}(n)|^2 \qquad (15)$$

where $\hat{x}(n)$ and $\hat{y}(n)$ are the known transmitted symbols in the training phase, and estimated symbols in the tracking phase; the estimation is performed either by comparing $x_o(n)$ (and $y_o(n)$) with thresholds or applying a maximum-likelihood (ML) estimation algorithm, as in the common 16QAM detector. The equalizer updating rule is again that of eqn. (7), but

the gradients are

$$\begin{aligned}
\nabla_{\mathbf{h}_{xx}^{(n)}} f_{LMS}(\mathbf{h}^{(n)}) &= C_{1,LMS} \cdot \left(\mathbf{x}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yy}^{(n)}} f_{LMS}(\mathbf{h}^{(n)}) &= C_{2,LMS} \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{xy}^{(n)}} f_{LMS}(\mathbf{h}^{(n)}) &= C_{1,LMS} \cdot \left(\mathbf{y}_i^{(n)}\right)^* \\
\nabla_{\mathbf{h}_{yx}^{(n)}} f_{LMS}(\mathbf{h}^{(n)}) &= C_{2,LMS} \cdot \left(\mathbf{x}_i^{(n)}\right)^*
\end{aligned} \tag{16}$$

which is again similar to eq. (8), but the coefficients $C_{1,LMS}$ and $C_{2,LMS}$ are

$$\begin{bmatrix} C_{1,LMS} \\ C_{2,LMS} \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} x_o(n) \\ y_o(n) \end{bmatrix} \tag{17}$$

$$\begin{aligned}
a &= 2\left(x_o(n) - \hat{x}(n)\right) \\
b &= 2\left(y_o(n) - \hat{y}(n)\right)
\end{aligned} \tag{18}$$

The LMS equalizer compares the output of the equalizer $x_o(n)$ directly to $\hat{x}(n)$, and $\hat{x}(n)$ is correct only if the phase noise is not too strong. Otherwise, a system which employs the LMS equalizer needs a phase synchronizer inside the loop which updates the equalizer coefficients. If the system correctly converges, then the LMS equalizer produces, the outputs $x_o(n) = x(n - L_x)e^{j\theta_x(n)}$ and $y_o(n) = y(n - L_y)e^{j\theta_y(n)}$, where $L_x = L_y$ and $\theta_x(n) = \theta_y(n) = 0$. It is still possible, however, that the degeneracy problem occurs, i.e. $x_o(n) = y_o(n)$.

The Stokes equalizer described in Sect. III is insensitive to phase noise and frequency offset as the CMA equalizer, but generates $x_o(n) = x(n - L_x)e^{j\theta_x(n)}$ and $y_o(n) = y(n - L_y)e^{j\theta_y(n)}$, where $L_x = L_y$ and $\theta_x(n) = \theta_y(n) = \theta(n)$; since the phase offsets are forced to be equal, only one phase synchronizer is needed which estimates $\theta(n)$ from the two inputs $x_o(n)$ and $y_o(n)$, with a signal to carrier noise which is twice that available with the CMA system. The degeneracy problem can occur also in this case.

V. Decision rule in the Stokes space

During the tracking phase, a decision $\hat{\mathbf{S}}(n)$ has to be made in the Stokes space on the transmitted symbol Stokes vector $\mathbf{S}(n)$, based on the equalized Stokes vector $\mathbf{S}_e(n)$ corresponding to the signal samples $x_o(n)$ and $y_o(n)$ at the output of the equalizer. In the case of PM-16QAM, $\hat{\mathbf{S}}(n)$ must be chosen among the $M_s = 60$ points shown in Fig. 3.

The minimum distance rule, applied in common detectors, is the following:

$$\begin{aligned}
\hat{\mathbf{S}}(n) &= \min_k{}^{-1} \left\|\mathbf{S}_e(n) - \mathbf{S}_{(k)}\right\|^2 = \\
&= \min_k{}^{-1} \left[\|\mathbf{S}_{(k)}\|^2 - 2\|\mathbf{S}_{(k)}\|\|\mathbf{S}_e(n)\| \cos(\theta_{(k)})\right]
\end{aligned} \tag{19}$$

with $k = 1, \ldots, M_s$. $\|\mathbf{A}\| = \sqrt{A_1^2 + A_2^2 + A_3^2}$ is the norm of vector $\mathbf{A}$, and $\theta_{(k)}$ is the angle between the two vectors $\mathbf{S}_e(n)$ and $\mathbf{S}_{(k)}$, such that

$$\begin{aligned}
\mathbf{S}_e(n) \cdot \mathbf{S}_{(k)} &= \sum_{i=1}^{3} S_{i,(k)} S_{i,e}(n) = \\
&= \|\mathbf{S}_e(n)\|\|\mathbf{S}_{(k)}\| \cos(\theta_{(k)})
\end{aligned} \tag{20}$$

However, the minimum distance criterion is optimum only if vector $\mathbf{D}(n) = \mathbf{S}_e(n) - \mathbf{S}(n)$ is a zero-mean Gaussian random vector, and this is not the case in Stokes space, since the channel additive Gaussian noise is non-linearly processed in the evaluation of the Stokes parameters of $x_o(n)$ and $y_o(n)$. In the case of a simple AWGN channel, if the transmitted symbol is associated with the Stokes vector $\mathbf{S}_{(k)} = [S_{1,(k)}, S_{2,(k)}, S_{3,(k)}]$, then the conditional probability density function (pdf) of the Stokes vector $\mathbf{S} = [S_1, S_2, S_3]$ associated with the noisy received symbol is [12]:

$$\begin{aligned}
f_{\mathbf{S}|\mathbf{S}_{(k)}}(\mathbf{S}|\mathbf{S}_{(k)}) = \frac{1}{16\pi\|\mathbf{S}\|\sigma^2} e^{-\frac{\|\mathbf{S}_{(k)}\|+\|\mathbf{S}\|}{2\sigma^2}} \cdot \\
\cdot I_0\left(\frac{\sqrt{\|\mathbf{S}_{(k)}\|\ \|\mathbf{S}\|}}{\sigma^2} \cos(\theta_{(k)}/2)\right)
\end{aligned} \tag{21}$$

where $\theta_{(k)}$ is the angle between the two vectors $\mathbf{S}$ and $\mathbf{S}_{(k)}$. In eq. (21), $I_0(x)$ is the modified Bessel function of the first kind and order zero, whereas σ^2 is the noise variance that can be evaluated on the conventional scattering diagrams for each polarization. Then the decided Stokes vector, according to the maximum likelihood criterion, is

$$\begin{aligned}
&\hat{\mathbf{S}}(n) = \max_k{}^{-1} f_{\mathbf{S}|\mathbf{S}_{(k)}}(\mathbf{S}_e(n)|\mathbf{S}_{(k)}) = \\
&= \max_k{}^{-1} \left[-\tfrac{\|\mathbf{S}_{(k)}\|}{2\sigma^2} + \log I_0\left(\tfrac{\sqrt{\|\mathbf{S}_{(k)}\|\ \|\mathbf{S}_e(n)\|}}{\sigma^2} \cos(\theta_{(k)}/2)\right)\right]
\end{aligned} \tag{22}$$

The simplification in eq. (22) is obtained by removing the term $\exp[-\|\mathbf{S}_e(n)\|/(2\sigma^2)]/(16\pi\|\mathbf{S}_e(n)\|\sigma^2)$, common to all the values of k, and taking the logarithm of the remaining part of the conditional pdf. Moreover, it is possible to use the following asymptotic approximation

$$\log I_0(x) \simeq x \tag{23}$$

so that the decision rule can be simplified as follows:

$$\begin{aligned}
\hat{\mathbf{S}}(n) &\simeq \max_k{}^{-1} \left[-\frac{\|\mathbf{S}_{(k)}\|}{2\sigma^2} + \frac{\sqrt{\|\mathbf{S}_{(k)}\|\ \|\mathbf{S}_e(n)\|}}{\sigma^2} \cos\left(\frac{\theta_{(k)}}{2}\right)\right] \\
&= \min_k{}^{-1} \left[\|\mathbf{S}_{(k)}\| - 2\sqrt{\|\mathbf{S}_{(k)}\|\ \|\mathbf{S}_e(n)\|}\ \cos\left(\frac{\theta_{(k)}}{2}\right)\right].
\end{aligned} \tag{24}$$

Note the similarity between the minimum distance decision rule in eqn. (19) and the approximate optimum detection rule in eqn. (24).

VI. Performance analysis

The performance of the proposed equalizer was assessed in a single-channel PM-16QAM system at 32 GBaud, with raised-cosine spectrum with roll-off 0.1. In order to test the equalizer performance in the presence of both inter-symbol interference and a misalignment between the polarizations

(which could be particularly critical for an update rule performed in the Stokes space), a residual dispersion (i.e. not compensated for by the CD equalizer) equal to 250 ps/nm and a differential group-delay (DGD) of one symbol were imposed. No phase noise was included in this first set of simulations. The BER values were estimated through Monte-Carlo simulation (over 2^{16} symbols) for several combinations of DGD axis and state of polarization (SOP) at the input of the Rx, for a total of 896 cases, uniformly distributed in Stokes space. The results are shown in Fig. 6 in terms of BER vs. OSNR (measured over a reference bandwidth equal to 0.1 nm) for both the Stokes-space and CMA update rules, with $N = 31$ taps. The reported curves show the average performance. For all cases, the range of variation of BER values around the average performance for all the considered values of SOP and DGD axis is always in the range $\pm 10\%$. In both cases, the equalizer taps were initialized using a known training sequence, and then switched to decision-directed operation after 10,000 symbols. A different value of μ was used in the training and tracking phases and both were optimized at each OSNR value. Differently from [8], in which a fixed-threshold decision criterion was used in the final QAM detector, the results shown in Fig. 6 have been obtained using a maximum-likelihood decision criterion (for both the SSA and CMA cases). For the SSA either the minimum distance criterion or the new metric described in Section V has been used when making a decision in the Stokes space. For high values of OSNR, the two metrics give similar result, whilst for low values of OSNR (i.e. BER values higher than 10^{-2}), the minimum distance metric fails in correctly estimating the transmitted symbols and consequently the performance of the equalizer is significantly degraded. Comparing the performance of CMA and SSA with the new metrics, no significant difference can be seen in Fig. 6, with both of them being very close to the theoretical limit (less than 0.5-dB penalty).

One advantage of the SSA over CMA is its higher convergence speed, i.e. it requires a shorter training sequence in order to achieve a stable performance. In Fig. 7, the convergence speed of the two update algorithms is compared at the same OSNR (18 dB). Results are shown in terms of BER as a function of the training sequence length (in number of symbols). The BER values shown in the figure have been derived as an average over all values obtained for different polarization states at the input of the receiver, for a total of 128 measures for each point. As an example, with SSA 3,000 symbols are sufficient to get a BER value of $2 \cdot 10^{-2}$, while with CMA a number of training symbols as high as 13,000 is required.

The performance of the equalizer was tested also in the presence of phase noise. In both cases, a radius-directed Viterbi&Viterbi algorithm was inserted after the butterfly equalizer for carrier phase estimation [13]. In principle, there should be no performance difference. However, since the Stokes-space update guarantees that the two polarizations after equalization are perfectly aligned to each other in phase (which is not true when CMA is used), the phase error estimate can be obtained as the average of both polarizations, thus making the algorithm more robust to amplifier spontaneous emission

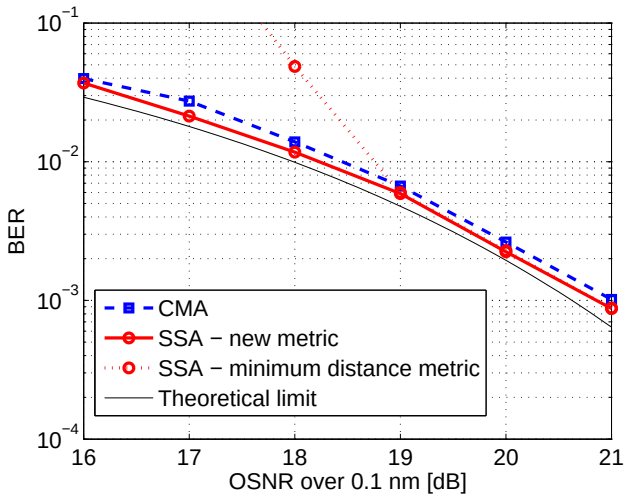

Fig. 6. BER vs. OSNR curve for buttery equalizer based on SSA and CMA update rules. The ideal theoretical curve is also shown as a reference.

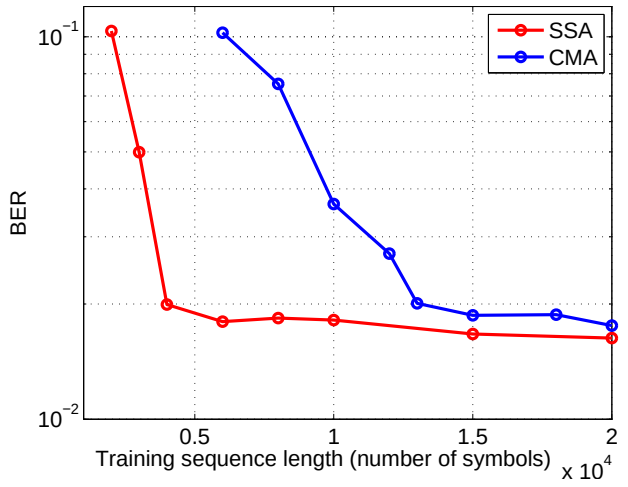

Fig. 7. BER vs. length of the training sequence (in number of symbols) for both SSA and CMA algorithm at OSNR=18 dB.

(ASE) noise and, as a result, more tolerant to phase noise. As an example, Fig. 8 shows the OSNR penalty (at a reference BER of $2 \cdot 10^{-2}$) with respect to the ideal performance without phase noise as a function of the combined linewidth $\Delta\nu$ of the Tx laser and local Rx oscillator (LO). The 1-dB penalty phase noise tolerance is 15 MHz with SSA and 9 MHz with CMA.

Finally, we also investigated the tolerance of the equalizers to a polarization-related effect like PDL, which in principle could be of harm to an equalizer based on an update rule in the Stokes space. We assumed to have a lumped PDL element before the insertion of ASE noise. While PDL on the signal is perfectly compensated by the butterfly equalizer, after equalization its effect is transferred to the noise components and its impact on performance depends on the PDL axis. We thus tested the performance by changing the amount of average PDL and the direction of the PDL axis (considering a total of

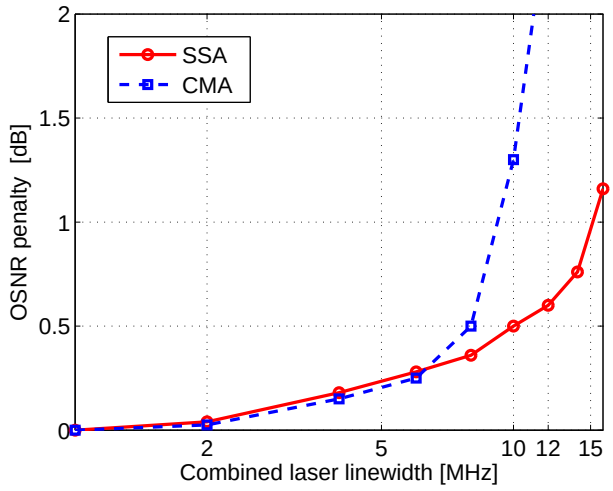

Fig. 8. OSNR penalty (at a reference BER equal to $2 \cdot 10^{-2}$) vs. the combined laser linewidth.

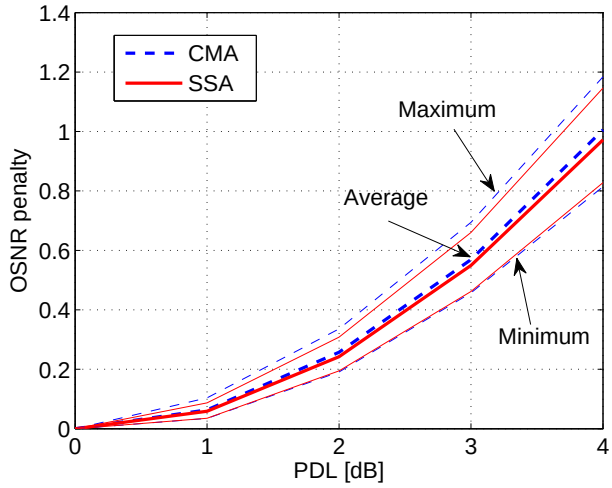

Fig. 9. OSNR penalty (at a reference BER equal to $2 \cdot 10^{-2}$) vs. average PDL.

500 PDL axes, uniformly distributed on the Poincaré sphere). The results are shown in Fig. 9 in terms of OSNR penalty (at a reference BER equal to $2 \cdot 10^{-2}$) with respect to the case without any PDL, as a function of the average PDL (in dB). Average, maximum and minimum values are shown in the figure. Interestingly, the SSA does not show any increased sensitivity to PDL with respect to CMA.

VII. Conclusions

In this paper we have shown a detailed analytical derivation of the equations used for performing the adaptive update of the butterfly equalizer in a coherent receiver, based on an error signal evaluated in Stokes space. We showed that the newly proposed SSA has a negligible increase in complexity with respect to commonly used CMA and LMS algorithms. With respect to LMS, it has the advantage of not requiring a feedback loop between the phase synchronizer and the butterfly equalizer. Moreover, it has the advantage of not being affected by the differential delay and differential phase problems typical of CMA: this fact enables the use of a CPE algorithm based on a phase error obtained as the average of both polarizations, which makes the algorithm more tolerant to phase noise. We also introduced a novel metric for taking a decision on the transmitted symbol in Stokes space, which turned out to significantly improve the performance with respect to the conventional minimum distance criterion, with a similar computational complexity. Finally, we tested the performance of the SSA in the presence of polarization-related effects like PMD and PDL, showing no increased sensitivity with respect to CMA.

VIII. Appendix

A. Evaluation of the gradients of the objective function

Assuming that $\mathbf{a}$ is a column vector with N complex known elements a_k and $\mathbf{h}$ is a column vector with N complex variables $e_k = e_{kr} + j e_{ki}$, with $k = 1, \ldots, N$, the following real function can be defined:

$$f_1(\mathbf{h}) = \Re\{\mathbf{a}^\dagger \mathbf{h}\} = \Re\{\mathbf{a}^* \cdot \mathbf{h}\} = \Re\left\{\sum_{k=1}^{N} a_k^* e_k\right\} = \\ = \frac{1}{2}\left\{\sum_{k=1}^{N} a_k^* e_k + \sum_{k=1}^{N} a_k e_k^*\right\} \quad (25)$$

where $\mathbf{a}^\dagger$ is the transpose and complex conjugate of $\mathbf{a}$. The derivative of $f_1(\mathbf{h})$ with respect to e_{nr} and with respect to e_{ni} is:

$$\frac{\partial}{\partial e_{nr}} f_1(\mathbf{h}) = \frac{1}{2}\{a_n^* + a_n\} = \Re\{a_n\}, \\ \frac{\partial}{\partial e_{ni}} f_1(\mathbf{h}) = \frac{1}{2j}\{a_n - a_n^*\} = \Im\{a_n\} \quad (26)$$

Then the derivative with respect to e_n is defined as

$$\frac{\partial}{\partial e_n} f_1(\mathbf{h}) = \frac{\partial}{\partial e_{nr}} f_1(\mathbf{h}) + j \frac{\partial}{\partial e_{ni}} f_1(\mathbf{h}) = \\ = \Re\{a_n\} + j\Im\{a_n\} = a_n \quad (27)$$

and the gradient of

$$f_1(\mathbf{h}) = \Re\{\mathbf{a}^\dagger \mathbf{h}\}$$

with respect to the column vector $\mathbf{h}$ is

$$\nabla_{\mathbf{h}} f_1(\mathbf{h}) = \mathbf{a} \quad (28)$$

Similarly, the gradient of

$$f_2(\mathbf{h}) = \Im\{\mathbf{a}^\dagger \mathbf{h}\}$$

is

$$\nabla_{\mathbf{h}} f_2(\mathbf{h}) = j\mathbf{a} \quad (29)$$

As a consequence, for the function

$$f_3(\mathbf{h}) = |\mathbf{a}^\dagger \mathbf{h}|^2 = \left(\Re\{\mathbf{a}^\dagger \mathbf{h}\}\right)^2 + \left(\Im\{\mathbf{a}^\dagger \mathbf{h}\}\right)^2 = \\ = f_1^2(\mathbf{h}) + f_2^2(\mathbf{h}) \quad (30)$$

the gradient is

$$\nabla_{\mathbf{h}} f_3(\mathbf{h}) = 2\left[\mathbf{a}^\dagger \mathbf{h}\right]\mathbf{a}. \quad (31)$$

Since $\mathbf{a}^\dagger\mathbf{h}$ is just a complex number, the above expression can also be written as

$$\nabla_{\mathbf{h}} f_3(\mathbf{h}) = 2\mathbf{a}\left[\mathbf{a}^\dagger\mathbf{h}\right] = 2\left[\mathbf{a}\mathbf{a}^\dagger\right]\mathbf{h}$$

where $\mathbf{a}\mathbf{a}^\dagger$ is a complex $N \times N$ hermitian matrix.

The objective function $f(\mathbf{h}(n))$ in eqn. (6) is

$$\begin{aligned} f(\mathbf{h}(n)) &= f(\mathbf{h}_{xx}(n), \mathbf{h}_{xy}(n), \mathbf{h}_{yx}(n), \mathbf{h}_{yy}(n)) = \\ &= (S_{1e}(n) - \hat{S}_1(n))^2 + (S_{2e}(n) - \hat{S}_2(n))^2 + \\ &+ (S_{3e}(n) - \hat{S}_3(n))^2 \end{aligned} \tag{32}$$

and

$$\begin{aligned} \nabla_{\mathbf{h}_{pq}} f(\mathbf{h}) &= 2(S_{1e}(n) - \hat{S}_1(n))\nabla_{\mathbf{h}_{pq}} S_{1e}(n) \\ &+ 2(S_{2e}(n) - \hat{S}_2(n))\nabla_{\mathbf{h}_{pq}} S_{2e}(n) + \\ &+ 2(S_{3e}(n) - \hat{S}_3(n))\nabla_{\mathbf{h}_{pq}} S_{3e}(n) \end{aligned} \tag{33}$$

where p and q can be both either x or y. Then we are left with the task of evaluating the gradient with respect to $\mathbf{h}_{xx}, \mathbf{h}_{yy}, \mathbf{h}_{xy}, \mathbf{h}_{yx}$ of each of the Stokes parameters $S_{1e}(n), S_{2e}(n), S_{3e}(n)$.

In particular, applying the rules (28), (29) and (31), we have

$$\begin{aligned} \nabla_{\mathbf{h}_{xx}} S_{1e}(n) &= 2(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{x}_i + 2(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{x}_i = 2x_o(n)\mathbf{x}_i \\ \nabla_{\mathbf{h}_{yy}} S_{1e}(n) &= -2(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{y}_i - 2(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{y}_i = -2y_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{xy}} S_{1e}(n) &= 2(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{y}_i + 2(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{y}_i = 2x_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{yx}} S_{1e}(n) &= -2(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{x}_i - 2(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{x}_i = -2y_o(n)\mathbf{x}_i \end{aligned} \tag{34}$$

$$\begin{aligned} \nabla_{\mathbf{h}_{xx}} S_{2e}(n) &= 2(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{x}_i + 2(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{x}_i = 2y_o(n)\mathbf{x}_i \\ \nabla_{\mathbf{h}_{yy}} S_{2e}(n) &= 2(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{y}_i + 2(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{y}_i = 2x_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{xy}} S_{2e}(n) &= 2(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{y}_i + 2(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{y}_i = 2y_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{yx}} S_{2e}(n) &= 2(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{x}_i + 2(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{x}_i = 2x_o(n)\mathbf{x}_i \end{aligned} \tag{35}$$

$$\begin{aligned} \nabla_{\mathbf{h}_{xx}} S_{3e}(n) &= 2j(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{x}_i + 2j(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{x}_i = 2jy_o(n)\mathbf{x}_i \\ \nabla_{\mathbf{h}_{yy}} S_{3e}(n) &= -2j(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{y}_i - 2j(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{y}_i = -2jx_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{xy}} S_{3e}(n) &= 2j(\mathbf{x}_i^\dagger\mathbf{h}_{yx})\mathbf{y}_i + 2j(\mathbf{y}_i^\dagger\mathbf{h}_{yy})\mathbf{y}_i = 2jy_o(n)\mathbf{y}_i \\ \nabla_{\mathbf{h}_{yx}} S_{3e}(n) &= -2j(\mathbf{x}_i^\dagger\mathbf{h}_{xx})\mathbf{x}_i - 2j(\mathbf{y}_i^\dagger\mathbf{h}_{xy})\mathbf{x}_i = -2jx_o(n)\mathbf{x}_i \end{aligned} \tag{36}$$

Combining (33)-(36), we obtain the final expression of the four gradients, shown in (37) below, and used in (8).

$$\begin{aligned} \nabla_{\mathbf{h}_{xx}} f(\mathbf{h}) &= 4\left[(S_{1e}(n) - \hat{S}_1(n))x_o(n)\right. \\ &\left. + (S_{2e}(n) - \hat{S}_2(n))y_o(n) + j(S_{3e}(n) - \hat{S}_3(n))y_o(n)\right]\mathbf{x}_i \\ \nabla_{\mathbf{h}_{yy}} f(\mathbf{h}) &= 4\left[-(S_{1e}(n) - \hat{S}_1(n))y_o(n)\right. \\ &\left. + (S_{2e}(n) - \hat{S}_2(n))x_o(n) - j(S_{3e}(n) - \hat{S}_3(n))x_o(n)\right]\mathbf{y}_i \\ \nabla_{\mathbf{h}_{xy}} f(\mathbf{h}) &= 4\left[(S_{1e}(n) - \hat{S}_1(n))x_o(n)\right. \\ &\left. + (S_{2e}(n) - \hat{S}_2(n))y_o(n) + j(S_{3e}(n) - \hat{S}_3(n))y_o(n)\right]\mathbf{y}_i \\ \nabla_{\mathbf{h}_{yx}} f(\mathbf{h}) &= 4\left[-(S_{1e}(n) - \hat{S}_1(n))y_o(n)\right. \\ &\left. + (S_{2e}(n) - \hat{S}_2(n))x_o(n) - j(S_{3e}(n) - \hat{S}_3(n))x_o(n)\right]\mathbf{x}_i \end{aligned} \tag{37}$$

REFERENCES

[1] S.J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804-817, 2008.

[2] P.J. Winzer, "High-spectral-efficiency optical modulation formats," *J. Lightw. Technol.*, vol. 30, no.24, pp. 3824-3835, Dec. 15, 2012.

[3] K. Kikuchi, S. Kawakami "Multi-level signaling in the Stokes space and its application to large-capacity optical communications," *Opt. Exp.*, vol. 22, no. 7, pp. 7374-7387, 2014.

[4] B. Szafraniec, B. Nebendahl, and T. Marshall, "Polarization demultiplexing in Stokes space," *Opt. Exp.*, vol. 18, no. 17, pp. 1792817939, 2010.

[5] Zhenming Yu, Xingwen Yi, Qi Yang, Ming Luo, Jing Zhang, Lei Chen, and Kun Qiu, "Experimental Demonstration of Polarization Demultiplexing in Stokes Space for Coherent Optical PDM-OFDM," in Proc. OFC 2013, paper OW3B.6, Anaheim, Mar. 2013.

[6] Zhenming Yu, Xingwen Yi, Qi Yang, Ming Luo, Jing Zhang, Lei Chen, and Kun Qiu, "Polarization demultiplexing in stokes space for coherent optical PDM-OFDM," *Opt. Exp.*, vol. 21, no. 3, pp. 3885, 2013.

[7] Di Che, An Li, Xi Chen, Qian Hu, Yifei Wang, and William Shieh, "160-Gb/s Stokes Vector Direct Detection for Short Reach Optical Communication," in Proc. OFC 2014, paper Th5C.7, San Francisco (USA), Mar. 2014.

[8] G. Bosco, M. Visintin, P. Poggiolini and F. Forghieri, "A Novel Update Algorithm in Stokes Space for Adaptive Equalization in Coherent Receivers," in Proc. of OFC 2014, paper Th3E.6, San Francisco (USA), Mar. 2014.

[9] G. Bosco, M. Visintin, P. Poggiolini, A. Nespola, M. Huchard and F. Forghieri, "Experimental Demonstration of a Novel Update Algorithm in Stokes Space for Adaptive Equalization in Coherent Receivers," in Proc. of ECOC 2014, paper Tu.3.3.6, Cannes (France), Sep. 2014.

[10] A. Beling, N. Ebel, A. Matiss, G. Unterbrsch, M. Nlle, J.K. Fischer, J. Hilt, L. Molle, C. Schubert, F. Verluise and L.Fulop, "Fully-integrated polarization-diversity coherent receiver module for 100G DP-QPSK, in Proc. OFC/NFOEC 2011, paper OML5, Los Angeles (USA), Mar. 2011.

[11] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ, Prentice-Hall, 1986.

[12] S. Benedetto and P. Poggiolini, "Multilevel polarization shift keying: optimum receiver structure and performance evaluation," *IEEE Trans. on Comm.*, vol. 42, no. 2/3/4, p. 1172, 1994.

[13] I. Fatadin and S. Savory., "Laser linewidth tolerance for 16QAM coherent optical systems using QPSK partitioning," *IEEE Photon. Technol. Lett*, vol. 22, no. 9, pp. 631-633, Nov. 15, 2010.