# Convergence Comparison of CMA and ICA for Blind Polarization Demultiplexing of QPSK and 16-QAM Signals

Pontus Johannisson(1), Henk Wymeersch(2), Martin Sjödin(1), A. Serdar Tan(2), Erik Agrell(2), Peter Andrekson(1), Magnus Karlsson(1)

(1) Photonics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-412 96 Göteborg, Sweden, ✉ pontus.johannisson@chalmers.se
(2) Communication Systems Group, Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

***Abstract*** *Two algorithms for blind polarization demultiplexing, based on the constant modulus criterion and independent component analysis, are compared. It is found that the latter converges significantly faster to within a given SNR penalty tolerance.*

## Introduction

The data rate can be doubled by using dual-polarization transmission in systems with coherent detection. Typically, polarization diversity is then used in the receiver and polarization demultiplexing (POLDEMUX) is performed through digital signal processing (DSP). The constant modulus algorithm (CMA)[1] is often used for POLDEMUX, but has the drawback that the two polarization tributaries can converge to the same channel, requiring ad-hoc solutions[2,3]. Furthermore, CMA is not designed for non-constant modulus formats such as 16 quadrature amplitude modulation (16-QAM). Independent component analysis (ICA) is a more rigorous alternative to CMA and relies on the assumption that the two channels are statistically independent. Although ICA has been suggested for POLDEMUX, very few results have been published[4]. The objective of both CMA and ICA is to quickly estimate the polarization state. This allows the system to switch to decision-directed mode[5] and the convergence time should be made short, in particular in packet-switched networks where rerouting is expected.

In this paper we make a direct comparison of the convergence properties of CMA and ICA for POLDEMUX. We find that ICA is capable of better estimating the polarization state from a given set of data within a given processing time.

## Problem formulation

We assume that chromatic dispersion has been compensated for in prior receiver components and that polarization-mode dispersion and polarization-dependent losses are negligible. Hence, we consider a system with polarization mixing, phase noise, and ASE noise. An equivalent block diagram is shown in Fig. 1. At time $k$,

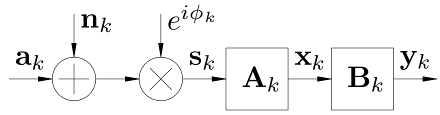


**Fig. 1**: System model showing the addition of noise, the phase drift, and the polarization change before the signal is sampled as $\mathbf{x}$.

the independent and identically distributed (i.i.d.) complex data symbols on the two polarizations, $\mathbf{a}_k$, are affected by complex additive white Gaussian noise (AWGN), $\mathbf{n}_k$, and an unknown phase rotation, $\phi_k$. The symbol phases need not be i.i.d. and can have an arbitrary distribution, possibly including symbol-by-symbol phase shifts from self-phase modulation. The sampled output is written $\mathbf{x}_k = \mathbf{A}_k\mathbf{s}_k$, where $\mathbf{s}_k = (s_k^{(\mathrm{X})}, s_k^{(\mathrm{Y})})^{\mathrm{T}}$ contains the two polarizations of the signal and $^{\mathrm{T}}$ denotes transposition. The complex $2\times 2$ matrix $\mathbf{A}_k$ is modeled as unitary, i.e., $\mathbf{A}_k^{\mathrm{H}}\mathbf{A}_k = \mathbf{I}$ and static over the observation time. Here, $^{\mathrm{H}}$ denotes Hermitian conjugation and $\mathbf{I}$ is the identity matrix.

The goal of both CMA and ICA is to quickly converge so that $\mathbf{y}_k = \mathbf{B}_k\mathbf{x}_k$ is a good estimate of $\mathbf{s}_k$. This estimate of $\mathbf{B}_k$ is based on the observations $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k$.

## Algorithm description

In CMA[1], we solve the optimization problem

$$\hat{\mathbf{B}}_{\mathrm{CMA}} = \arg\min_{\mathbf{B}} \mathbb{E}\left[\sum_{m\in\{\mathrm{X},\mathrm{Y}\}} \left(|y_k^{(m)}|^2 - \rho^2\right)^2\right], \qquad (1)$$

where $\mathbb{E}[\cdot]$ is the expectation operator and $\rho^2 = \mathbb{E}[|s^{(\mathrm{X})}|^4]/\mathbb{E}[|s^{(\mathrm{X})}|^2] = \mathbb{E}[|s^{(\mathrm{Y})}|^4]/\mathbb{E}[|s^{(\mathrm{Y})}|^2]$. The stochastic gradient descent update rule is

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \mu\boldsymbol{\psi}(\mathbf{y}_k)\mathbf{x}_k^{\mathrm{H}}, \qquad (2)$$

where $\mu > 0$ is a step size parameter and $\boldsymbol{\psi} = (\psi^{(\mathrm{X})}, \psi^{(\mathrm{Y})})^{\mathrm{T}}$ with $\psi^{(m)}(\mathbf{y}_k) = (|y_k^{(m)}|^2 - \rho^2) y_k^{(m)}$, $m \in \{\mathrm{X}, \mathrm{Y}\}$.

In ICA[6], we aim to find $\mathbf{B}$ such that the distribution of $\mathbf{y}_k$ is close to the marginal distribution of $\mathbf{s}_k$, averaged over all possible realizations of $\phi_k$. Here, *closeness* is measured in terms of the Kullback-Leibler divergence between both distributions, leading to the optimization problem

$$\hat{\mathbf{B}}_{\mathrm{ICA}} = \arg\min_{\mathbf{B}} \int p_{\mathbf{Y}}(\mathbf{s}) \log \frac{p_{\mathbf{Y}}(\mathbf{s})}{p_{\mathbf{S}}(\mathbf{s})} \mathrm{d}\mathbf{s}, \quad (3)$$

where the integration occurs over $\mathbb{C}^2$. The corresponding stochastic gradient minimization rule is found to be[6]

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \mu \Bigg[ \underbrace{\frac{1}{p_{\mathbf{S}}(\mathbf{y}_k)} \frac{\partial p_{\mathbf{S}}(\mathbf{y}_k)}{\partial \mathbf{s}^*}}_{\equiv \boldsymbol{\varphi}(\mathbf{y}_k)} \mathbf{y}_k^{\mathrm{H}} - \mathbf{I} \Bigg] \mathbf{B}_k. \quad (4)$$

In the derivation we have used the approach called *the relative gradient descent*[6] that does not require inversion of $\mathbf{B}_k$. An alternative update rule when $\mathbb{E}(\mathbf{s}_k \mathbf{s}_k^{\mathrm{H}}) = \mathbf{I}$ is the *orthogonal update*[6]

$$\begin{aligned}\mathbf{B}_{k+1} = \mathbf{B}_k - \mu[\boldsymbol{\varphi}(\mathbf{y}_k)\mathbf{y}_k^{\mathrm{H}} - \mathbf{y}_k \boldsymbol{\varphi}(\mathbf{y}_k)^{\mathrm{H}} \\ + \mathbf{y}_k \mathbf{y}_k^{\mathrm{H}} - \mathbf{I}]\mathbf{B}_k. \quad (5)\end{aligned}$$

To perform the update of $\mathbf{B}$ in ICA, the PDF $p_{\mathbf{S}}(\mathbf{s})$ needs to be known. The values $s^{(\mathrm{X})}$ and $s^{(\mathrm{Y})}$ are not independent, due to the common phase, but we approximate $p_{\mathbf{S}}(\mathbf{s}) = p_S(s^{(\mathrm{X})}) \times p_S(s^{(\mathrm{Y})})$. Now, $s^{(m)}$ has a uniform phase and an amplitude that is a mixture of Ricean PDFs. At typical medium-to-high SNR values, we can approximate the PDF $p_S(\cdot)$ as

$$p_S(s^{(m)}) = \sum_a \frac{1}{D_a} \exp\left[ -\frac{(|s^{(m)}| - |a|)^2}{N_0} \right], \quad (6)$$

where $m \in \{\mathrm{X}, \mathrm{Y}\}$, $D_a = D_a(N_0, |a|)$ is a normalization constant, and the summation occurs over the set of symbols in the constellation.

## Numerical simulation results

Numerical simulations have been performed to compare the convergence rate of CMA and ICA. For ICA the update rule in (5) has been used together with the approximate PDF in (6). The $\mathbf{A}$ matrix is drawn uniformly from the set of $2 \times 2$ unitary matrices, and is then held constant during the simulation. This is reasonable for limited observation times, say, 2000 symbols, as this corresponds to 0.2 $\mu$s at 10 Gbaud, which is too short a time to have any significant changes of the polarization state. The initial estimate of $\mathbf{B}$ is set to $\mathbf{B}_0 = \mathbf{I}$. A symbol sequence is generated and complex AWGN noise is added, corresponding to a nominal BER of $10^{-3}$. To measure the convergence performance, we compute the SNR degradation as follows: Let $E_{\mathrm{s}}$ be the average energy per symbol, and $N_0$ the noise PSD so that the nominal SNR is $E_{\mathrm{s}}/N_0$. Introducing $\mathbf{C} = \mathbf{B}_k \mathbf{A}$, the measured SNR at time $k$ on polarization X is

$$\mathrm{SNR}_k^{(\mathrm{X})} = \frac{|C_{1,1}|^2 E_{\mathrm{s}}}{|C_{1,2}|^2 E_{\mathrm{s}} + N_0(|C_{1,1}|^2 + |C_{1,2}|^2)}. \quad (7)$$

Calculating $\mathrm{SNR}_k^{(\mathrm{Y})}$ in a similar way, the SNR penalty at time $k$ is

$$\mathrm{SNR}_k^{\mathrm{pen}} = E_{\mathrm{s}}/N_0 - \min[\mathrm{SNR}_k^{(\mathrm{X})}, \mathrm{SNR}_k^{(\mathrm{Y})}]. \quad (8)$$

Running a large number of simulations using different $\mathbf{A}$ matrices we can compute the probability of being below a given SNR penalty threshold at every iteration. This value is then used as a measure of the convergence rate and we have set the convergence threshold value to 1 dB SNR penalty. The step sizes $\mu$ have been selected to maximize the final probability of being below 1 dB penalty.

Figs. 2 and 3 show results for QPSK and 16-QAM, respectively. The thick lines correspond to symbol-by-symbol updates (2) and (5). The thin lines show cumulative updates where at time $k$ we perform a gradient descent based on all observations up to time $k$. This cumulative approach has higher computational complexity but shows the achievable performance gains by making use of all available information at every time step.

For the *symbol-by-symbol* updates, ICA corresponds to the blue line (marked with circles) and CMA to the two red lines. The solid line for CMA corresponds to an implementation suggested by Kikuchi[2], orthogonalizing the rows of $\mathbf{B}_k$ at every time step. It is seen that ICA outperforms CMA for both modulation formats since it has a significantly lower probability of being above the SNR penalty threshold. This shows that ICA is much more efficient in estimating the polarization state from a given set of data. However, the probability for being below 1 dB penalty is above $10^{-3}$ for both algorithms also after 2000 symbols in the 16-QAM case. For ICA this is because local minima in the problem (3) slow down the convergence. We expect that a more careful selection of $\mathbf{B}_0$ would improve the situation.

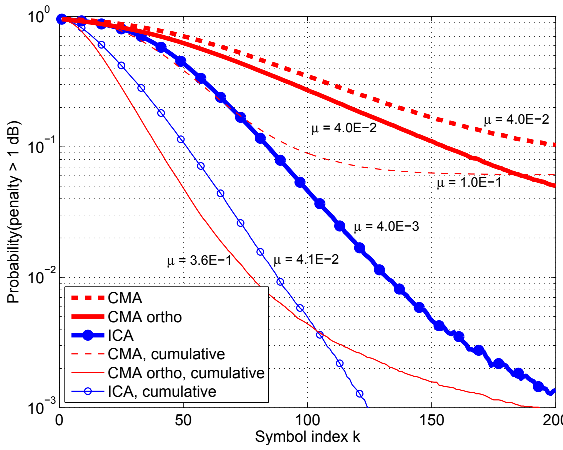


**Fig. 2**: Probability for being below 1 dB penalty for CMA and ICA on 200 symbols of QPSK data. The step size, $\mu$, is indicated close to the different curves.

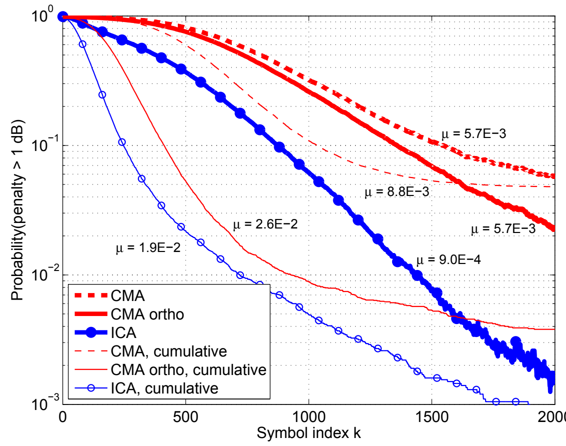


**Fig. 3**: Probability for being below 1 dB penalty for CMA and ICA on 2000 symbols of 16-QAM data. The step size, $\mu$, is indicated close to the different curves.

For the *cumulative* updates, the simulation results are shown using thin lines in Figs. 2 and 3. Particularly striking is the improvement for CMA in the QPSK case, showing that basing the gradient estimate on more than one symbol has the potential of speeding up the convergence rate considerably.

## Computational complexity

CMA is widely used due to its simplicity and low computational complexity. For 16-QAM the described ICA algorithm is quite costly (without approximations), but for QPSK the difference from CMA is small. To see this we write out the two corresponding expressions explicitly.

As seen from (2) the updating for CMA is done according to $\mathbf{B}_{k+1} = \mathbf{B}_k + \mathbf{M}_k$, where the update matrix $\mathbf{M}_k$ has the elements

$$M_{1,1} = -\mu(|y_1|^2 - \rho^2)y_1x_1^*, \tag{9}$$
$$M_{1,2} = -\mu(|y_1|^2 - \rho^2)y_1x_2^*, \tag{10}$$
$$M_{2,1} = -\mu(|y_2|^2 - \rho^2)y_2x_1^*, \tag{11}$$
$$M_{2,2} = -\mu(|y_2|^2 - \rho^2)y_2x_2^*, \tag{12}$$

where $y_1 = y_k^{(X)}$ etc for brevity. Using ICA as described by (5) and (6), the update is multiplicative according to $\mathbf{B}_{k+1} = \mathbf{N}_k\mathbf{B}_k$, where $\mathbf{N}_k$ has the elements

$$N_{1,1} = 1 + \mu(1 - |y_1|^2), \tag{13}$$
$$N_{1,2} = \frac{\mu|a|}{2\sigma^2}(e^{i\phi_1}y_2^* - y_1e^{-i\phi_2}) - \mu y_1y_2^*, \tag{14}$$
$$N_{2,1} = \frac{\mu|a|}{2\sigma^2}(y_1^*e^{i\phi_2} - e^{-i\phi_1}y_2) - \mu y_1^*y_2, \tag{15}$$
$$N_{2,2} = 1 + \mu(1 - |y_2|^2), \tag{16}$$

with the phases $\phi_1 = \measuredangle y_1$ and $\phi_2 = \measuredangle y_2$. Counting the multiplications we see that (9)–(12) need 16 and (13)–(16) need 14. In addition, ICA needs 2 phase extractions and 8 multiplications to compute $\mathbf{N}_k\mathbf{B}_k$. How much difference this makes in practise depends on the exact hardware implementation, but we see that the extra computational effort in ICA is very limited.

## Conclusions

We have compared the constant modulus algorithm and independent component analysis for POLDEMUX in terms of their convergence speed. It was found that, for a given number of symbols and a set SNR penalty limit, ICA has a significantly lower probability of failure. The difference is more than a factor of ten and the increase in computational effort is small.


## Acknowledgement

We acknowledge the financial support from the Swedish Governmental Agency for Innovation Systems (VINNOVA) and the EU EURO-FOS project.


## References


1 D. N. Godard, IEEE Trans. Commun., **28**, 1867 (1980).
2 K. Kikuchi, IEEE/LEOSST 2008, MC2.2 (2008).
3 L. Liu et al., OFC 2009, OMT2, (2009).
4 H. Zhang et al., ECOC 2008, Mo.3.D.5, (2008).
5 S. J. Savory, Opt. Express, **16**, 804 (2008).
6 J.-F. Cardoso, Proc. IEEE, **86**, 2009 (1998).