Master Thesis

Digital Nonlinear Pre-Distortion for Coherent Optical Transceivers

Santiago Cañas Casco

Supervisors:
Dr.-Ing. Johannes Fischer
Prof.-Ing. Andrea Carena

at

Fraunhofer Institute for Telecommunications,
Heinrich Hertz Institute, HHI
Santiago Cañas Casco

29.06.2018

Statement

I assure that this thesis was written by me and in my own words, except for quotations from published sources which are indicated and acknowledged as such. I am conscious that the addition of material from other works or a paraphrase of such material without acknowledgement will be treated as plagiarism, according to the Politecnico di Torino and HHI Conduct of Examinations.

29. Jun 2018

Santiago Cañas Casco
Müllerstraße, 34
Berlin 13353
santiago.canas.casco@hhi-extern.fraunhofer.de

Acknowledgements

I would like to thank my advisors Johannes Fisher and Andrea Carena for guiding and supporting me over this months. They have set an example of excellence as a researcher and instructor.

I would like to thank all my collegues at HHI, research engineers, collaborators, and multitude of graduates who contributed to this research. I am very grateful to all of you. I would like to specially thank my dear collegues Luzt Molle and Carsten Schmidt-Langhorst for their constant enthusiasm and support in the enhancement of this thesis.

I would especially like to thank my amazing family for the love, support, and constant encouragement. In particular, I would like to thank my parents. I could not have done this without you.

Abstract

In order to satisfy the continually increasing demand for more capacity in fiber-optical networks, spectrally efficient modulation formats such as quadrature amplitude modulation with high cardinality have to be employed.

However, non-ideal characteristics of optical transmitter components such as digital-to-analog converter, modulator driver amplifier and electro-optic modulator cause distortions of the generated optical waveforms leading to a reduced quality of transmission.

One promising option to compensate these distortions are digital nonlinear pre-distortion filters applied in the transmitter digital signal processing unit. The goal of the thesis is to develop advanced regularization algorithm in penalized least squares approach for transmitter impairment estimation in order to determine the optimum pre-distortion filters. The algorithms are implemented and verified in Matlab. Finally, the implemented algorithms were validated in laboratory experiments.

List of abbreviations

ACO	Analog Coherent Optics
ADC	Analog-to-digital Converter
ASIC	Application specific integrated circuits
AWG	Arrayed Waveguide Grating
CD	Chromatic Dispersion
CFR	Carrier Frequency Recovery
CPE	Carrier Phase Estimation
CPR	Carrier Phase Recovery
CV	Cross - Validation
DA	Driver - Amplifier
DAC	Digital-to-Analog Converter
DP - IQ Modulator	Dual - Polarization IQ Modulator
DSP	Digital Signal Processing
EVM	Error Vector Magnitude
k_i	kernel coefficient of order i
LED	Light Emitting Diodes
LO	Local Oscillator
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MZM	Mach - Zehnder Modulator
PBC	Polarization Beam Combiner
PBS	Polarization Beam Splitter
PDM	Pol Division Multiplexing
PLS	Penalized Least Square Error
PMD	Polarization Mode Dispersion
SNR	Signal to noise ratio
SOP	State of Polarization
GCV	Generalized - Cross - Validation
OLS	Ordinary Least Square Error
VLC	Visible Light Communication

List of Figures

Table of Contents

Chapter 1

Introduction

1.1 Motivation and outline contributions

The attainable information rates in optical systems have been extensively increased over the past decades with the launch of optical amplifiers, coherent detection, modulation formats and digital signal processing techniques.

The recovery of signal amplitude and signal phase using coherent optical detection constitutes a compelling information for compensation and effective mitigation of the transmission impairments. This coherent detection establishes a technique for optical communication to push the performance closer to the Shannon capacity limit.

The long-haul high-capacity performance in optical communication systems is significantly deteriorated by transmission impairments such as polarization mode dispersion, phase noise and nonlinearities among others. Polarization mode dispersion can be adaptively equalized. Phase noise can be estimated and compensated using carrier phase estimation. Nonlinearities can be commonly mitigated by using pre-distortion at the receiver.

Pre-distortion is a compensation method for optical communication systems. In pre-distortion, the idea is to add a compensation module preceding the channel it is about to compensate. The pre-distorter is a nonlinear inverse filter which aims to counteract the nonlinear characteristics of the channel so that the merged block of these two is almost linear.

Notice the difference between pre-distorter and equalizer, the latter is used to reverse the distortions drawn by a signal transmitted through a channel. Pre-distortion can be interpreted as a design process for adapting the signal to the nonlinearities of the channel assuming knowledge of the channel at the transmitter.

Pre-distorters can be implemented using model structures such as Volterra series, parallel Hammerstein structures, or lookup tables (LUT) among others. The model estimation can be seen as a least squares problem, which minimize the sum of squared residuals. The residuals are the difference between an observed value and the fitted value provided by the model.

The odd comes when the least squares approach deals with short data (more predictors than observations) or in low SNR regimes. The primary focus of the work presented in this thesis is to examine an alternate path to the least-squares solution under bad conditions. The alternative is a regularization algorithm to the penalized least squares (PLS) solution, a regularization function which introduces additional *a priori* information of the signal to the estimation.

This study bring us one step closer to solving the problem of nonlinear signal processing, and how it can be better used to obtain accurate estimates in noisy environments.

1.2 Organization of the thesis

The master thesis is divided into six chapters.

Chapter I : Introduction

In chapter I of the thesis, an introduction is briefly shown where the motivation and organization of the thesis are presented.

Chapter II : Application Scenarios

Chapter II presents different scenarios where the digital nonlinear pre - distortion can be applied with significant improvement in performance. Due to the number of nonlinear components in a coherent system, a coherent communication system is a suitable scenario for the digital nonlinear pre - distortion implementation.

Chapter III : Theoretical Approach

In chapter III, a theoretical approach discussion about nonlinear systems is shown. Volterra kernels are discussed and the different methods or techniques to obtain them, such as ordinary least squares error (OLS) and penalized least squares error (PLS). In the latter case, the regularization solution proposed by this thesis is discussed as well as regularization parameter λ.

Chapter IV : Simulation Results

Chapter IV introduces the different results obtained by a set of simulations implemented using Matlab. This chapter constitutes a set of simulations to test the preformance of the algorithms implemented in this thesis.

Chapter V : Experimental Results

Chapter V discussess the different results obtained by the actual experiment setup implemented at the HHI lab.

Chapter VI : Conclusion

First, chapter VI presents the actual problem with the proposed solution of the thesis. Then, the procedure used in order to validate the performance of the implemented algorithms is explained. Finally, a summary and dissertation of the results are presented and conclusions are drawn from the empirical results along with different directions towards future research.

Chapter 2

Application Scenarios

2.1 Coherent System

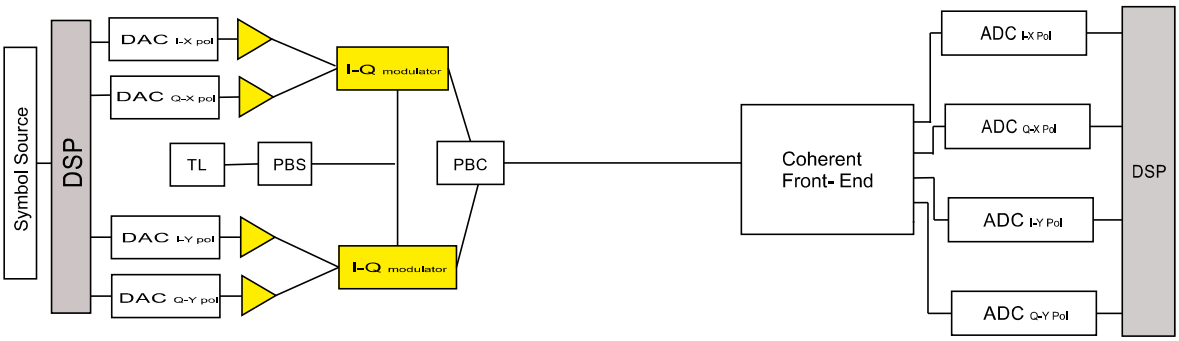


Fig 2.1: Scheme of a coherent system.

It is a well known fact that the market of coherent optical communications systems employing polarization division multiplexing (PDM) and coherent detection assisted by digital signal processing (DSP) contributed to the evolution of long-haul transmission. The computation is solved by optoelectronic modules with high-speed, application-specific integrated circuits (ASICs) in coherent optical receivers.

This solution empowers adaptive electronic equalization of linear transmission impairments such as polarization mode dispersion (PMD) and polarization - dependent loss (PDL). This solution allows to compensate all linean effects at the analog optical transmitter and at the receiver front-end, such as time skew of quadrature components along with quadrature and polarization imbalance.

Finally, all the digital receiver functionalities such as digital clock recovery, intermediate frequency offset and phase-noise estimation can be performed.

Let us present the basic structure of a coherent system with a brief description of each of its parts.

2.1.1 Optical Transmitter: Optical Modulator

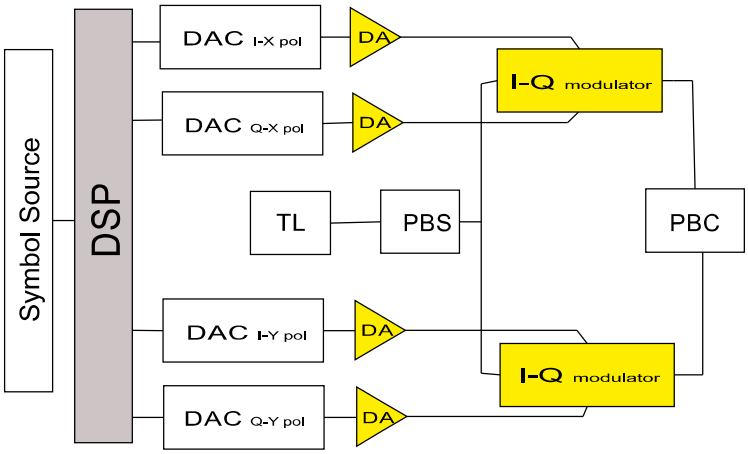


Fig 2.2: Scheme of a optical transmitter.

The first part of any communication system is the transmitter module. In figure 2.2, there are four different paths with different delay which leads to different phase noise from the structure digital - to - analog converter (DAC) + driver amplifier (DA) + IQ modulator. The aim is to estimate these effects caused by the previous structure and compensate them in the coherent transmitter.

Two outputs from the DAC + DA structure feed each I-Q modulator. The I-Q modulator box represents an amplitude/phase modulation in which a Super Mach-Zehnder modulator is used to encode one data stream into an optical carrier for the components I and Q. This approach is also known as Super Mach-Zehnder structure.

The lightbeam from the laser is split by the polarization beam splitter (PBS). Finally, the two polarization $\mathbf{X}$ and $\mathbf{Y}$ are combined by using a polarization beam combiner (PBC).

In the previous coherent system scheme figure 2.2, the different components of the optical transmitter marked in yellow correspond to those components responsible for adding nonlinear effects to the system. These are the components:

- ***Driver Amplifier (DA)*** Amplifiers usually operate as a linear device under small amplitude signal conditions and become nonlinear and distorting with the increasing drive level and the resulting distortion. Nonlinearity in amplifiers means that the output signal power does not vary in direct proportion to the input signal power [3].
- ***Mach-Zehnder Modulator (MZM)*** since MZMs have a nonlinear (sinusoidal) characteristic response it is a major cause of nonlinear effects. In practice, the linear regime is achieved by narrowing the driving signals to a linear interval of the MZMs response but limiting the system capacity in term of OSNR penalties for nonlinear distortions or by using nonlinear pre-distortion as a linearization method.

2.1.2 Transmission fiber

The second part of the coherent transmission system previously proposed is the channel, optical fiber.

- ***Silica Fibers*** is a dielectric medium which, in presence of an electric field induces electric charges within the material, creating a dielectric polarization field. The induced field will have an orientation such that the equivalent field within the dielectric declines. Therefore, the interaction between the electrical field and dielectric produces a modulation of the equivalent field within the medium.

Such modulation is a nonlinear function of the input field power. This phenomenon is the cause of the Kerr effect, which appears in silica fibers and causes nonlinear effects in optical communication systems. The Kerr effect can be seen as a change in the refractive index of the material proportional to the incident electrical power field [4][5].

2.1.3 Optical Receiver: Coherent Detection

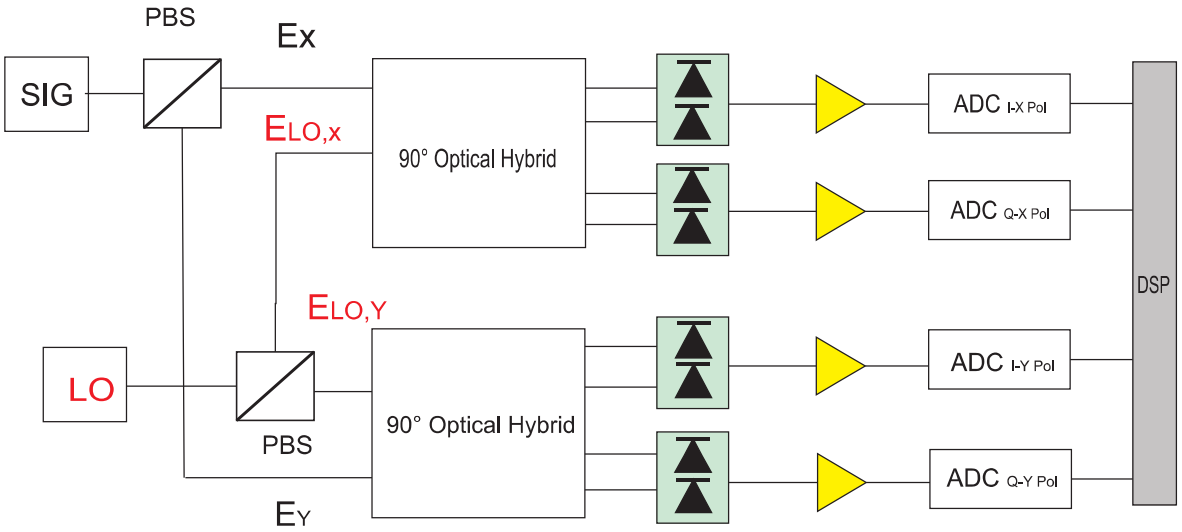


Fig 2.3: Scheme of the coherent receiver.

The last part of the coherent transmission system is the coherent detection. Coherent detection offers flexibility in the modulation formats, as the information can be encoded in amplitude and phase, or in both in-phase (I) and quadrature (Q) components of the carrier.

Coherent detection requires the knowledge of the transmitted carrier phase at the receiver: the received signal is demodulated by a local oscillator (LO) that is used as an absolute phase reference [6].

The figure 2.3 shows a schematic of a standard coherent optical front - end. The optical carrier and the local oscillator (LO), after passing through the polarization beam splitters are directed towards two optical 90 degrees hybrids that act as an IQ demodulator and it is used to mix the signal with the local oscillator (LO) as well as with the 90 degree shifted version of the local oscillator (LO). Then, there are balanced photo - detectors marked in green in the figure 2.3, after the detection of the outputs in the balanced photodiodes, the **in-phase** and **quadrature** components of the data signal (referenced to the continuous wave CW local oscillator) are restored.

The coherent receiver is employed to detect the two orthogonal polarizations of the received signal (the polarizations are separated using a polarization beam splitter) [6][7].

2.2 Thesis Application Scenario

As it has been described in the preceding sections, the sources of nonlinear effects are wide and distributed all along the coherent system. The main aim of this thesis is to design and implement a regularization method for the pre-distortion of nonlinear effects caused by the transmission module (DAC + DA + IQ mod). The pre-distortion process takes place in the coherent transmitter.

Chapter 3

Theoretical Approach

In this chapter we introduce the theoretical approach to pre-distortion, this is the schematic to be applied at the Kernels estimation module in the coherent detection system with pre-distortion. This pre - distortion estimates the transmitter nonlinear effect at the receiver level of the system.

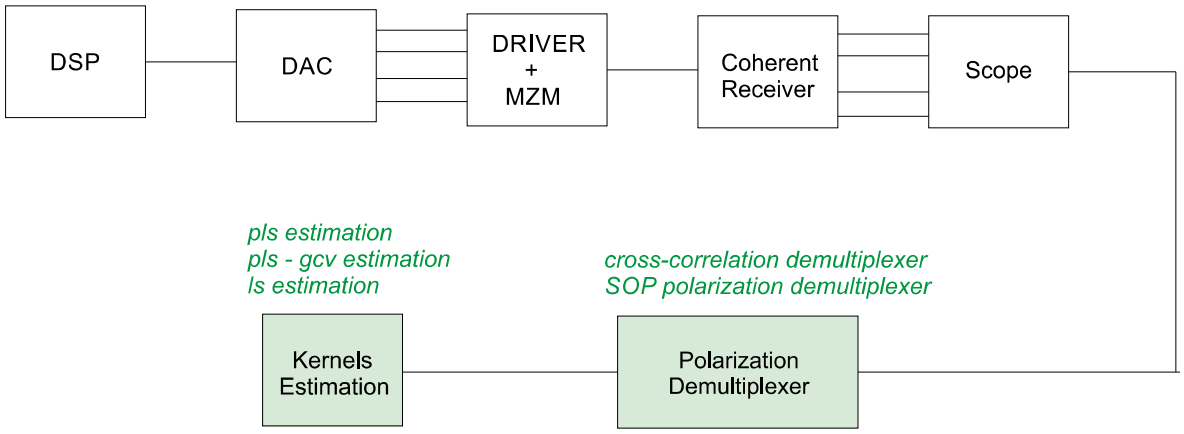


Fig 3.1: Brief thesis schematics.

The first part of the scheme consists on a simple coherent system. The received signal at the oscilloscope (scope) is processed by a typical digital signal processing dsp chain with a polarization demultiplexer.

At this point, in the polarization demultiplexing, there are two approaches, polarization demultiplexing based on cross-correlation or polarization demultiplexing based on state of polarization SOP. The difference in performance is discussed in the simulation chapter.

Finally, the kernels estimation module which compensate the nonlinear effects of the transmitter can use different algorithms; ordinary least squares (ols), penalized least squares (pls) and penalized least squares with cross - validation (pls - gcv). The regularization function in the penalized least squares is the major goal of this thesis.

Let us have a theoretical introduction for a better understanding of the kernel estimation.

3.1 Nonlinear Systems

Nonlinear systems are very common, and have been studied in the science field. To model and analyze nonlinear systems and solve related problems, people have carried out extensive studies, and developed a collection of mathematical theories, among which the Volterra series is one of the most famous.

Volterra series guaranteed to explain nonlinear behaviors, but the computational process was complex and it could only be applied to the analysis of simple nonlinear systems. This problem limited its application to a small range of practical engineering issues but it computation was very slow. This situation changed in the 1990s with the promotion of computer technology [11].

3.2 Volterra Series

Volterra series is one of the early approaches to achieve a characterization of a nonlinear system. It is a powerful mathematical tool for nonlinear system analysis. Actually, it is an extension of the standard convolution description of linear systems to nonlinear systems.

The Volterra series is a model for nonlinear behavior similar to the Taylor series [12] [13]. It can be interpreted as a generalization of the Taylor series extension for the analytic functions.

The Taylor series response of a nonlinear system depends only in the input at that particular time. Nevertheless, in the Volterra series, the response of the nonlinear system depends on the input to the system at all times.

Volterra Series

If a system is linear and time-invariant, then the linear input-output relation of the system can be represented by the convolution integral, which is shown as follows;

$$y(t) = \int h(t-\tau)x(\tau)d\tau \tag{3.1}$$

where, $y(t)$ is the output and $x(t)$ is the input and the system is determined uniquely by the impulse response function $h(t)$.

In contrast, for nonlinear continuous time-variant systems with memory, if the level of energy of the input signal x(t) is limited, the system response can be described by Volterra series.

$$y(t) = h_0 + \int h_1(\tau)x(t-\tau)d\tau + ... \iint h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 \tag{3.2}$$

where, h_1, h_2,..., h_n are the different orders Volterra kernel functions, which are expansions of the impulse response for the linear system to the nonlinear system. Then, the term h_0 represents the DC component of the system.

Volterra system can also be considered in the discrete domain:

$$H_p\{y[n]\} = \sum_{k_1=0}^{m_{p-1}} ... \sum_{k_p=k_{p-1}}^{m_p-1} h_p[k_1,...,k_p]y[n-k_1-\tau_1]...y[n-k_p-\tau_p] \tag{3.3}$$

Let us consider the following example with a third order memory $m = [m_1, m_2, m_3]$ considering the order dependent delay τ_p constant.

$$H_1\{y[n]\} = \sum_{k_1=0}^{m_1-1} h_1[k_1].y[n-k_1]. \tag{3.4}$$

$$H_2\{y[n]\} = \sum_{k_1=0}^{m_1-1} ... \sum_{k_2=k_1}^{m_2-1} h_2[k_1,k_2]y[n-k_1]y[n-k_2]. \tag{3.5}$$

$$H_3\{y[n]\} = \sum_{k_1=0}^{m_3-1} ... \sum_{k_3=k_2}^{m_3-1} h_3[k_1,k_2,k_3]y[n-k_1]y[n-k_2]y[n-k_3] \tag{3.6}$$

The Volterra series is fully determined by the kernels h_p, where h_1, h_2, h_3 stand for the first, second, and third order Volterra kernels [12]. The memory lengths are represented by m_p, where the index p refers to the order of the kernel system[12].

3.3 Volterra Kernels Identification

The estimation of Volterra kernels is a difficult problem. By its inverse nature, is a linear inverse problem since the system model must be determined from input/output measurements from the system [11] [13]. It does not assure at least one of the three assumptions of a well-posed problem (1. existing solution, 2. convergence to a unique solution, 3. the solution changes continuously with respect to the initial conditions). This fact brings instability in the inversion process.

Typically, the kernels representation needs a large number of parameters, with the number increasing geometrically with the order of the kernel. The first-order kernel represents the linear dynamics of the system while the higher-order kernels describe the nonlinear dynamics. It should be clear that, for a linear system, the first-order kernel is equivalent to the impulse response of the system. Therefore, the Volterra theory can be viewed as an expansion of the concept of linear convolution to nonlinear systems.

3.3.1 Ordinary Least Squares (OLS)

The aim of this linear regression approach is to minimize the sum of the squares of the residuals, residuals are the difference between the fitted value by the model and the observation.

On one hand, the positive aspect of this approach is that it is simple. On the other hand, the negative aspect of this approach is that it does not achieve the highest performance under a noisy environment and/or short data set [13].

$$\mathbf{X}\theta = \mathbf{Y} \qquad \theta \in \Re^k \qquad \mathbf{Y} \in \Re^n \tag{3.7}$$

$$\mathbf{X}^\top(\mathbf{X}\theta - \mathbf{Y}) = 0 \quad \rightarrow \quad \mathbf{X}^\top\mathbf{X}\theta - \mathbf{X}^\top\theta = 0 \tag{3.8}$$

$$\hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.9}$$

$$\hat{\theta}_{\mathrm{OLS}} = \arg\min_{\theta} \sum_{i=1}^{n} \epsilon_i^2 = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \tag{3.10}$$

Let us consider the case of an arbitrary nonlinear channel estimation. The following proposed channel consists on an exponential decay along the different kernel coefficients.

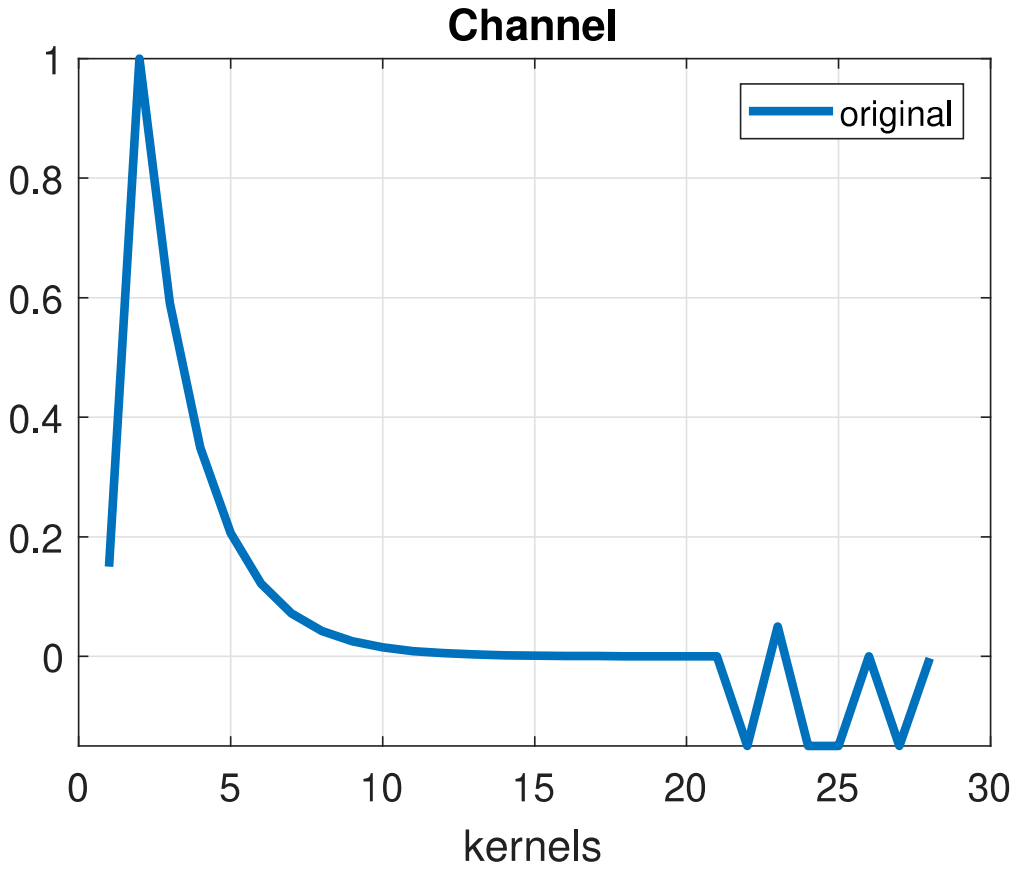


Fig 3.2: Original channel.

The following figures 3.3, 3.4 and 3.5 represent the normalized kernel coefficients for a Volterra memory $m = [m_1, m_2, m_3] =$ **[20, 3, 1]**, a **SNR = 30 dB** and a result **MSE = 4,59E-05**.

The setup of the simulation consists in generating a signal with an arbitrary filter shape. Then, the signal goes through the channel it is intended to estimate in the simulation. Finally, the generated input signal, the output signal (through channel) and the Volterra memory are used to estimate the channel.

Simulation setup:
- filter shape: **'rootraisedcosine'**
- filter order: **3**
- filter roll-off: **0.1**
- filter bandwidth: **0.5 GHz**
- Volterra memory: [**20, 3, 1**] - SNR: **30 dB**
- error parameter: **MSE**

Mean squared error is defined as:

$$MSE = (1/n)\sum(Y - \hat{Y})^2$$

Where $\hat{Y}$ corresponds to the predicted values with the kernel estimation, $\mathbf{Y}$ corresponds to the observation and n represents the number of samples.

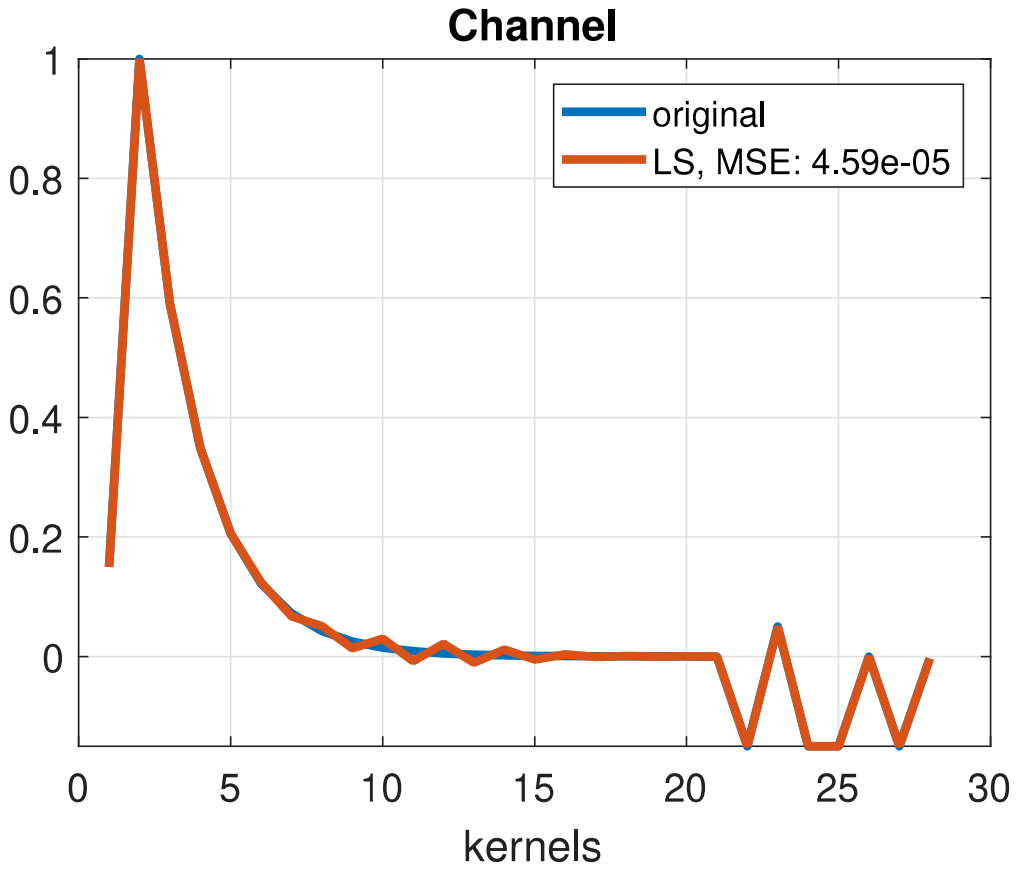


Fig 3.3: Least squares estimation of the arbitrary nonlinear channel.

The following picture 3.4 describes the kernel visualization for the **Ix** component estimation.

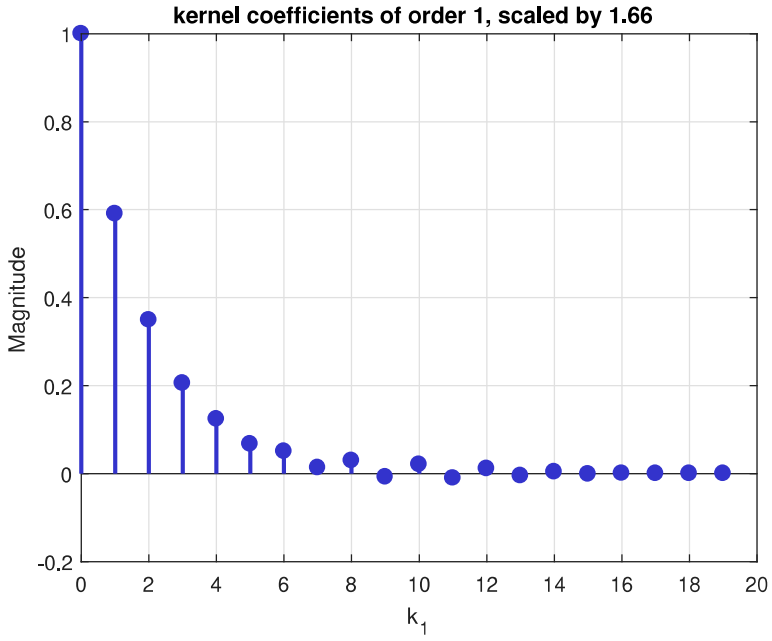


Fig 3.4: kernel coefficients of first order.

The linear component corresponds to the channel to be estimated.

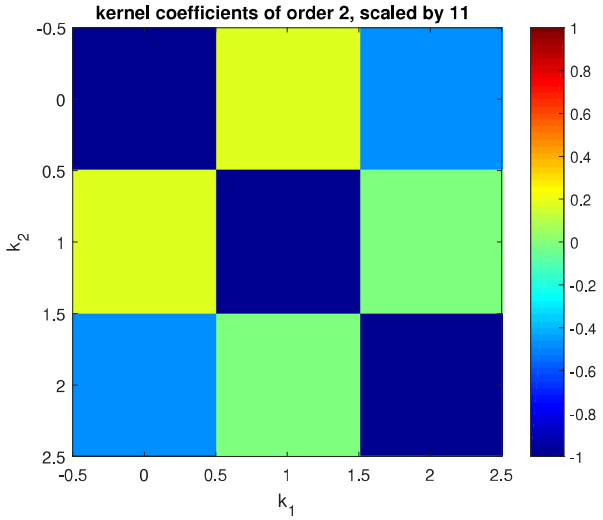


Fig 3.5: kernel coefficients of second order.

3.3.2 Penalized Least Squares (PLS)

It is well known that polynomial regression models, such as the Volterra filter, often suffer from inverse linear problem. Therefore, least squares Volterra filter estimates obtained from short noisy data are often very bad in accuracy[13].

The least squares approach does not meet all the requirements of a well-posed problem, this is the reason why some regularization must be added to the least squares equation. This regularization or penalization function contains *priori* information of the signal (main contribution of this thesis).

Penalized least squares is a well-known method for regularizing the least squares estimator and simulations have shown major improvements in the Volterra filter estimates. The main contribution of this thesis is the design of appropriate penalizing functions for the Volterra filter estimation problem [13] [15].

$$y_i = \mathbf{x}_i^T \theta + \eta_i \qquad 1 \leqslant i \leqslant n \tag{3.11}$$

Where $\mathbf{x}_1$,...,$\mathbf{x}_n$ are known column vectors and η_1,...,η_n are scalar observation noises. Gathering the observations into a vector, we have the following description:

$$\mathbf{y} = \mathbf{X}\theta + \eta \tag{3.12}$$

The ordinary least squares (OLS) estimator is:

$$\hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.13}$$

And $\hat{\theta}$ minimizes the sum of squared errors

$$\hat{\theta}_{min} = \frac{1}{n}||\mathbf{Y} - \mathbf{X}^T\theta||^2 \tag{3.14}$$

The problem with the OLS estimator is that if the noise is large enough, then demanding $\hat{\theta}$ to adapt to the data, will produce a model that takes the noise as well as θ as a parameter. In other words, it starts modelling the noise. For that reason, the least squares estimator may result in a very large variance with short noisy data records [13].

An approach to address this issue is known as the method of regularization.

$$\hat{\theta}_{min} = \frac{1}{n}||\mathbf{Y} - \mathbf{X}^T\theta||^2 + \lambda J(\theta) \qquad \lambda > 0 \tag{3.15}$$

Where $J(\theta)$ is a non-negative real-valued function which measures the roughness of θ. Here, the estimation problem consists in two penalizations. The first term of the right side of the equation 3.15 represents a penalization that comes from the data, the second term of the right side of the equation 3.15 comes from prior information of the signal.

Penalized methods usually come with at least one regularization parameter λ that controls the trade off between likelihood (variance) and penalty (bias). In other words, the regularization parameter decides which penalization is going to have more relevance, the penalization referring to data or the penalization coming from prior information of the signal.

Notice that for $\lambda = 0$ there is no longer the penalized least squares (PLS) method but the ordinary least squares (OLS) method defined previously.

Thus, when this penalized function will be a trade off between fidelity to the data and smoothness of the function.

$$J(\theta) = \theta^T R\theta \qquad J(\theta) = \sum_{j=1}^{m} (\theta_{j-1} - 2\theta_j + \theta_{j+1})^2 \tag{3.16}$$

$$\hat{\theta}_{PLS}(\lambda) = (\mathbf{X}^T\mathbf{X} + n\lambda R)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.17}$$

3.4 Regularization Matrix R

The following section is a discussion about the different approaches towards this regularization matrix R in order to obtain an accurate estimation.

$$\hat{\theta}_{PLS}(\lambda) = (\mathbf{X}^T\mathbf{X} + n\lambda R)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.18}$$

3.4.1 Projection Matrix

The kernel-penalizing matrix is the projection onto the low - dimensional subspace. The projection matrixes used in this approach correspond to $\mathbf{T}_p$ and $\mathbf{U}_p$ [13].

$$\hat{\theta}_{PLS}(\lambda) = (\mathbf{X}^T\mathbf{X} + n\lambda\mathbf{U}_p^T\mathbf{R}_p\mathbf{U}_p)^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.19}$$

3.4.1.1 Derivation of $\mathbf{T}_p$ and $\mathbf{U}_p$

First, considering the set of all possible p - tuples, then g_k are the unique generating p - tuples. The matrix $\mathbf{T}_p$ is constructed as follows. The kth row of the $\mathbf{T}_p$ corresponds to the kth unique generating p - tuple $g_k = (i_{1,k}, ..., i_{p.k})$. Let n_k be the number of distinct permutation of g_k [13]

Let us study the case of computing the matrix T_p in relation with h_p which contains all the indexes of the Volterra Kernel of the p - order. Moreover θ_h contains the symmetric indexes of the Volterra Kernel coefficients. Following the notation in [13], let's verify that $\theta_h = T_p\ h_p$ for $m = 3$ and $p = 2$.

$$\left(h_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \end{bmatrix} \right) \left(\theta_h = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 1 \\ 1 & 2 \\ 2 & 2 \end{bmatrix} \right)$$

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

So the previous equation will result in: $\theta_h = T_2 \; h_2$

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 1 \\ 1 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \end{bmatrix} \tag{3.20}$$

Paying attention on how the second index is computed, it can be appreciated that:

$$\begin{bmatrix} 0 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{3.21}$$

Then the third index...

$$\begin{bmatrix} 0 & 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 0 & 2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 2 & 0 \end{bmatrix} \tag{3.22}$$

In the computation of the matrix T, the number of elements per row that are different from zero is given by the number of permutation of the elements of the index corresponding to that row, for instance:

The index [0 2] has two permutations because the elements of this index can be written as: [0 2] or [2 0]. While the index [1 1] has one permutation because it can only be

written like that.

The other part of the computation remains in how the columns of the elements different from zero are placed. The columns are the correspondent positions of the permutations of our index in the expanded version of all the Volterra indexes.

Let's see it for the case of the index [0 2] with permutations [0 2] and [2 0] . Searching for the permutations at h_2 we found the position **3** and **7**.

$$h_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ \mathbf{0} & \mathbf{2} \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \mathbf{2} & \mathbf{0} \\ 2 & 1 \\ 2 & 2 \end{bmatrix}$$

Those previous positions **3** and **7** will be the columns of the elements different from zero at the row of the index [0 2] in θ_h in the T matrix.

$$T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Columns 3 and 7 at the third row in this case.

3.4.1.2 Order of the vector h_p

Following the notation given by [13]

$$y(k) = h_p^T d^{(p)}(k) \tag{3.23}$$

Where $d^{(p)}(k)$ detones the p - fold tensor product of the vector d(k) with itself.

For example, if

$$d(k) = [x(k), x(k-1)]^T \tag{3.24}$$

and p = 2, then

$$d^{(2)}(k) = [x^2(k), x(k)x(k-1), x(k-1)x(k), x^2(k-1)]^T \tag{3.25}$$

The equivalent in matlab is the kronecker tensor product which is defined as [30]:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1N}B \\ a_{21}B & a_{22}B & \dots & a_{2N}B \\ \vdots & \vdots & \ddots & \\ a_{N1}B & a_{N2}B & \dots & a_{NN}B \end{bmatrix}$$

With A [N x N] matrix.

Let's try the following example with d(k) $= [x(k), x(k-1), x(k-2)]^T$ and $p = 3$ (third order). Let's consider x(k) $= [1, 2, 3]^T$

$$d(k) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

$d^{(2)}(k)$ = d(k) $\otimes$ d(k)

$$d^{(2)}(k) = \begin{bmatrix} \mathbf{1} & 2 & 3 & 2 & \mathbf{4} & 6 & 3 & 6 & \mathbf{9} \\ \mathbf{2} & 3 & 1 & 4 & \mathbf{6} & 2 & 6 & 9 & \mathbf{3} \\ \mathbf{3} & 1 & 2 & 6 & \mathbf{2} & 4 & 9 & 3 & \mathbf{6} \\ \mathbf{2} & 4 & 6 & 3 & \mathbf{6} & 9 & 1 & 2 & \mathbf{3} \\ \mathbf{6} & 2 & 4 & 9 & \mathbf{3} & 6 & 3 & 1 & \mathbf{2} \\ \mathbf{3} & 6 & 9 & 1 & \mathbf{2} & 3 & 2 & 4 & \mathbf{6} \\ \mathbf{6} & 9 & 3 & 2 & \mathbf{3} & 1 & 4 & 6 & \mathbf{2} \\ \mathbf{9} & 3 & 6 & 3 & \mathbf{1} & 2 & 6 & 2 & \mathbf{4} \end{bmatrix}$$

The columns marked in red corresponds to the $d^{(2)}(k)$ expansion shown below

$$d^{(2)}(k) = \begin{bmatrix} x^2(k) \\ x(k)x(k-1) \\ x(k)x(k-2) \\ x(k-1)x(k) \\ x^2(k-1) \\ x(k-1)x(k-2) \\ x(k-2)x(k) \\ x(k-2)x(k-1) \\ x^2(k-2) \end{bmatrix}$$

$$h_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}$$

So when the product $y(k) = h_2^T d^{(2)}(k)$ takes place, the combination of the products correspond with the index of the Volterra Kernel.

With $\mathbf{U} \approx \mathbf{T}^T$

$$U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let the matrix $\mathbf{S}$ denotes the orthogonal projector corresponding to the penalized subspace [13]. The Fourier series corresponds to orthogonal projections of a given function onto the trigonometric polynomials.

$$\mathbf{S} = \mathbf{F}^{-1}\mathbf{H}\mathbf{F} \tag{3.26}$$

Where $\mathbf{F}$ is the Fourier transform of the identity matrix and $\mathbf{H}$ is a diagonal matrix which corresponds to a dynamic filter matrix based on spectrum shape of the input signal or also called smooth term [13].

Each p - order has a different projection matrix $\mathbf{S}_{(p)}$. The matrix $\mathbf{S}_{(p)}$ will correspond to $\mathbf{S}_{(p)} = \mathbf{F}_{(p)}^{-1}\mathbf{H}_{(p)}\mathbf{F}_{(p)}$. In each case, $\mathbf{H}_{(p)}$ is a diagonal matrix with a dimension m_p, which is the memory associated to the p - order.

This means that the diagonal elements of $\mathbf{H}_{(p)}$ are the m_p samples of the spectrum shape of the input signal. This is the aspect of the regularization matrix $\mathbf{R}$ based of orthogonal projection:

$$\mathbf{U}_p^T[(\mathbf{I}-\mathbf{S})^{(p)}]\mathbf{U}_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{U}_1^T[(\mathbf{I}-\mathbf{S}_1)^{(1)}]\mathbf{U}_1 & 0 & 0 \\ 0 & 0 & \mathbf{U}_2^T[(\mathbf{I}-\mathbf{S}_2)^{(2)}]\mathbf{U}_2 & 0 \\ 0 & 0 & 0 & \mathbf{U}_3^T[(\mathbf{I}-\mathbf{S}_3)^{(3)}]\mathbf{U}_3 \end{bmatrix}$$

Where the first 1 in the left-top corner of the regularization matrix corresponds to the DC component of the system and $[(\mathbf{I}-\mathbf{S})^{(p)}]$ stands for the p - order kroneker expansion of $(\mathbf{I}-\mathbf{S})$ which corresponds to the non-penalizing subspace [13]. Two versions can be applied for the non-penalizing subspace, the previous one with $[(\mathbf{I}-\mathbf{S})^{(p)}]$ which penalizes the high frequency components (thesis approach) or the $[\mathbf{I}^{(p)} - \mathbf{S}^{(p)}]$ which penalizes the low frequency components [13].

In order to verify the performance of this approach a base band signal has been employed to test the system. While the considered spectrum is the root raised cosine filter, which means that the diagonal elements of $\mathbf{H}_{(p)}$ are the m_p samples of the spectrum shape of the input with $p = 1...P$ while P is the number of orders of the Volterra kernels.

The content of $\mathbf{H}_{(1)}$, $\mathbf{H}_{(2)}$ and $\mathbf{H}_{(3)}$ will differ depending on the m_p, memory of the order p

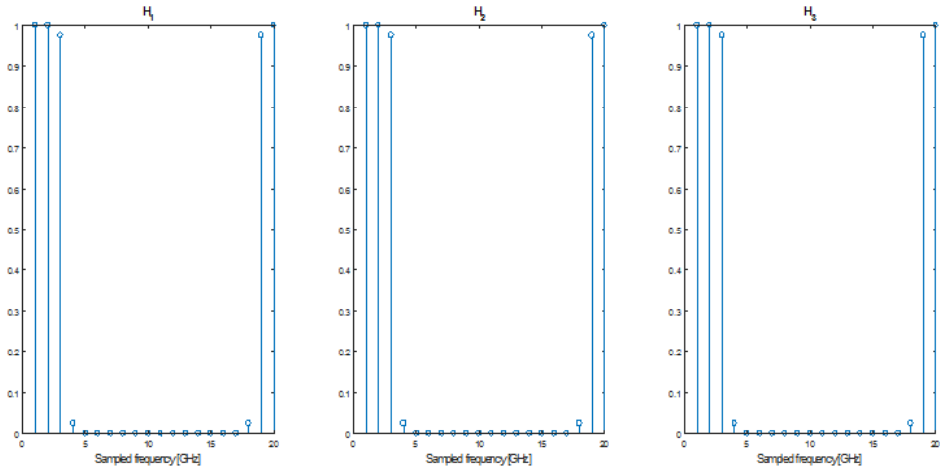


Fig 3.6: Sampled envelope of the H_p filter with $p = 1...3$ and $m_1 = m_2 = m_3 = 20$.

The frequency follows this scheme: reading from left to right the frequencies appreciated are; first the terms around DC component, then the positive frequencies and finally the negative frequencies. The filter applied to shape the input signal is a **root raised cosine** with a **roll - off: 0.25** and bandwidth = **8 GHz**.

3.5 Regulator parameter (λ)

Coming back to the equation 3.15

$$\hat{\theta}_{min} = \frac{1}{n}||\mathbf{Y} - \mathbf{X}^T\theta||^2 + \lambda J(\theta) \qquad \lambda > 0$$

The first term of the right side of the equation 3.15 represents a penalization that comes from the data, the second term of the right side of the equation 3.15 comes from prior information of the input signal.

The parameter λ plays a roll of trade-off on how much it is going to be emphasize one term with respect to the other.

3.5.1 Cross - Validation (CV)

Cross-validation assesses the fit of the model with a particular λ in a very similar way to the residual sum of squares, but removes the value y_i and observes how well the fit predicts the previously removed value [13] [14].

In other words, it attempts to minimize the residual sum of squares while ignoring the closest value; it is often referred to as the leave - one - out strategy. The problem with this model is that it assumes periodicity of the signal.

$$V(\lambda) = \frac{1}{n}\sum_{i=0}^{n}(y_i - x_i^T\hat{\theta}(\lambda))^2 \tag{3.27}$$

3.5.2 Generalized - Cross - Validation (GCV)

For the determination of the optimum smooth parameter, the thesis uses the generalized cross validation (GCV) approach. It is based on the principle of cross validation in which, the regularized inverse problem is solved by omitting at each realization one point of the observed data [13]. Then, the periodicity problem of the cross validation (CV) method is solved by adding different weights to the samples.

For each realization is used a different regularization parameter λ. It is expected that the adequate λ is chosen, that is, the estimated data parameters is near to the observed data parameters.

$$V_o(\lambda) = \frac{\frac{1}{n}\sum_{i=0}^{n}(y_i - x_i^T\hat{\theta}(\lambda))^2}{1 - \frac{t(\lambda)}{n}^2} \qquad t(\lambda) = \sum_{i=0}\frac{1}{1+\lambda k_i} \tag{3.28}$$

k_i are the eigenvalues of $(\mathbf{X}^T\mathbf{X})^{-1}R$

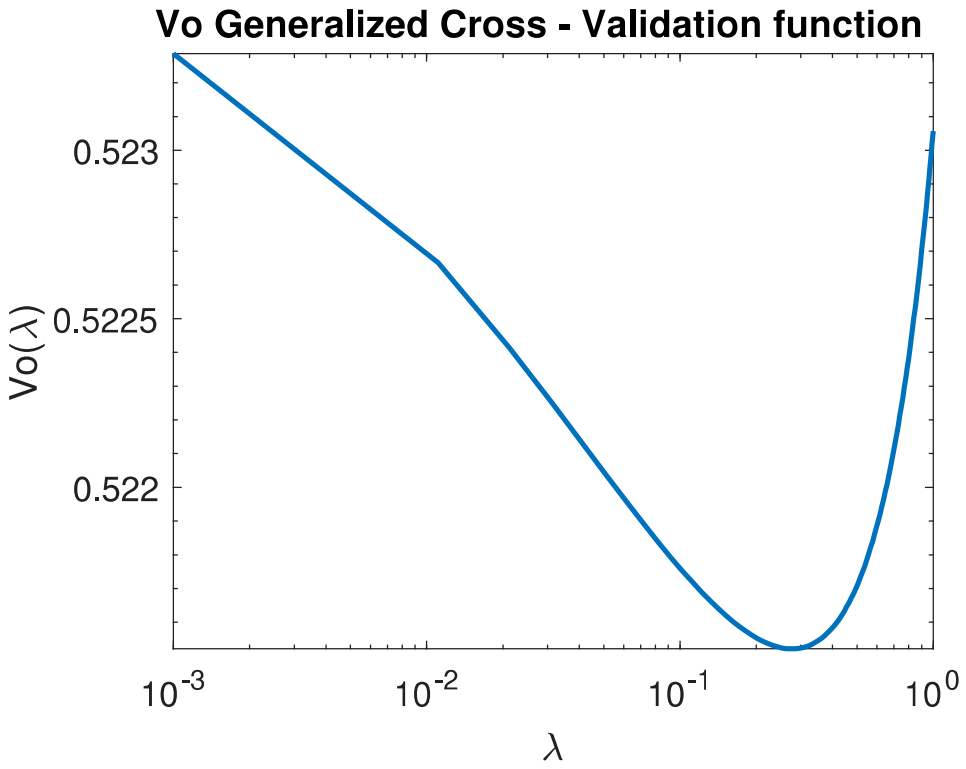


Fig 3.7: $V_o(\lambda)$ Generalized Cross - Validation Function.

In the previous illustration 3.7, the specific value of λ that minimizes the generalized cross validation function $V_o(\lambda)$ is the value $\lambda = 0.2$. On one hand, the region previous to the λ_{opt} it is known as under fitting because it is not the minimum error estimation. On the other hand, the region posterior to the λ_{opt} is known as overfitting and it gives us high level of error.

Let us see some comparative results between a prediction based on a generalized cross validation against a no regularization parameter predictor. The comparison is established on the mean squared error parameter for different values of SNR.

The setup of the simulation consists in generating a signal with an arbitrary filter shape. Then, the signal goes through the channel it is intended to estimate in the simulation. Finally, the generated input signal, the output signal (through channel) and the Volterra memory are used to estimate the channel while using generalized cross validation and no regularization parameter in the penalized least squares approach.

Setup of the measurement:
- number of samples: **1e5 samples**
- memory system: [**15, 3, 1**]
- signal-to-noise range: **-30 dB to 50 dB**
- spectrum: **OFF**
- filter shape: **ON**
    - filter shape = **'gauss'**
    - filter order = **3**
    - filter roll - off: **0.1**
    - filter bandwidth: **0.2 GHz**

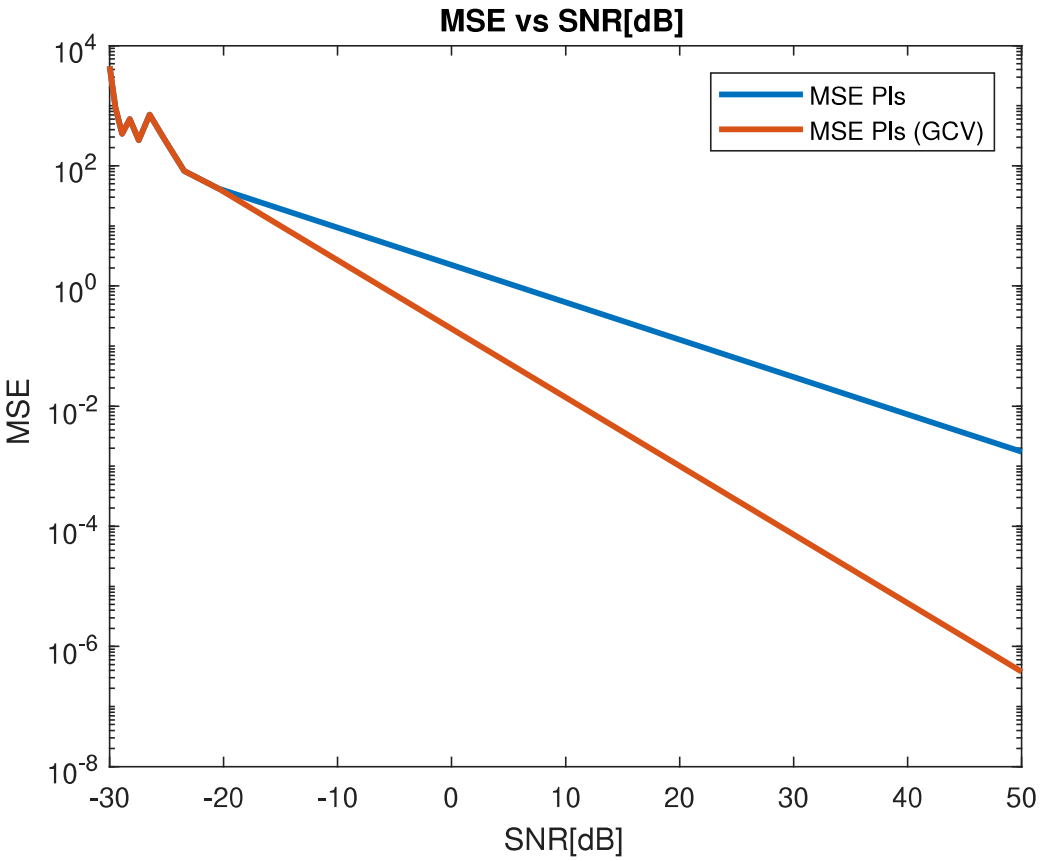


Fig 3.8: Comparison between a penalized least squares approach without a generalized cross validation and a penalized least squares approach with generalized cross validation implementation.

The result shows the same behavior for low values of SNR, after that, the performance of the penalized least squares approach with a generalized cross validation shows better performance.

Chapter 4

Simulation Results

This chapter discusses the different simulation results obtained in the performance test of the ordinary least squares (OLS) approach in comparison with the penalized least squares (PLS). Then, a performance test for the polarization demultiplexer with an acquired signal from the HHI laboratory is implemented.

The following simulations display the difference between the ordinay least squares (OLS) approach and the penalized least squares (PLS) approach, this difference becomes clear in the presence of a noisy environment and short data sets. The lower the signal to noise ratio SNR parameter, the greater it is the divergence in the accuracy while obtaining the correct kernel coefficients for both approaches. In general terms, the (PLS) approach provides a more accurate estimation than the (OLS) approach in low SNR regimes.

4.1 Simulation Setup

The simulation setup consists in a coherent transmission system constituted by interconnecting the transmitter and the receiver of the transceiver for analog coherent optics CFP2 - ACO property of the HHI laboratory.

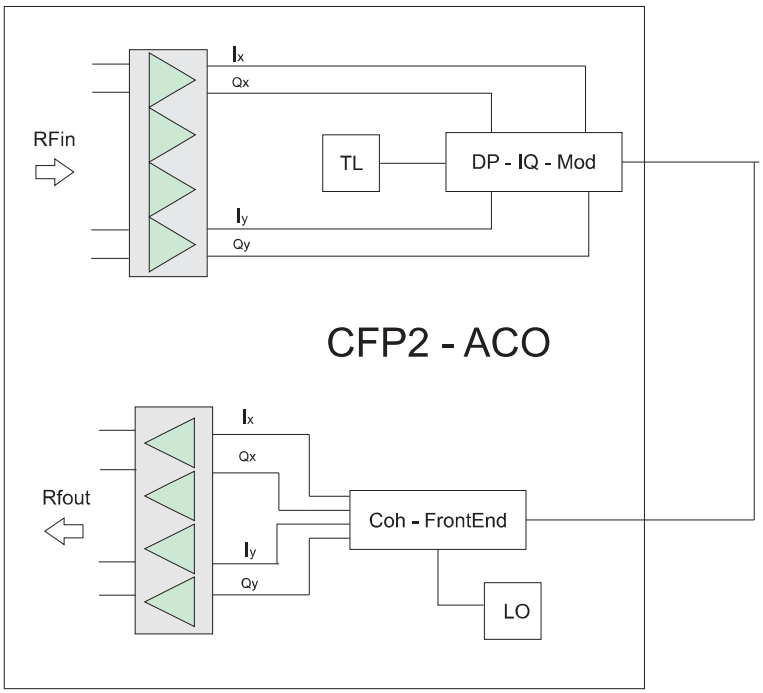


Fig 4.1: Scheme of the set - up to apply in the simulations

The transmitter, which is positioned in the top of the figure 4.1, consists in four radio frequency signals (I and Q components of both polarizations **X** and **Y**) entering an array of driver amplifiers marked in green at figure 4.1. Then, the different signal components (I_x, Q_x, I_y and Q_y) are coupled together at the dual polarization IQ modulator DP - IQ - Modulator.

The receiver, which is positioned at the bottom of the figure 4.1, consists on a coherent - front end module explained in section 2.1.3 to be finally decoulped into the four components to be amplify by an array of driver amplifiers marked in green in figure 4.1 at the receiver.

4.2 Dependency of PLS estimation with the bandwidth and roll-off

In order to check the performance of the PLS approach in relation to the bandwidth and the roll-off parameters of the input signal, a simulation test is proposed. Let us consider a filtered input signal by a raised cosine with a fixed value for the bandwidth and roll-off parameters. The vector h_p explained in the section 3.2 can be estimated as the sampled version of the filter shape used to filter the input of the system or as the spectrum shape calculated in relation to the autocorrelation function of the input signal of the system with itself. The transfer functions of each p-order H_p are explained in section 3.4.1.2 Let's see some illustrations to show the differences:

The setup of the simulation consists in:
- filter bandwidth: **0.5 GHz**
- filter shape: **'raisedcosine'**
- filter order: **3**
- roll-off: **0.1**
- SNR : **30 dB**
- physical center frequency: **193.1 THz / 1553.599 nm**

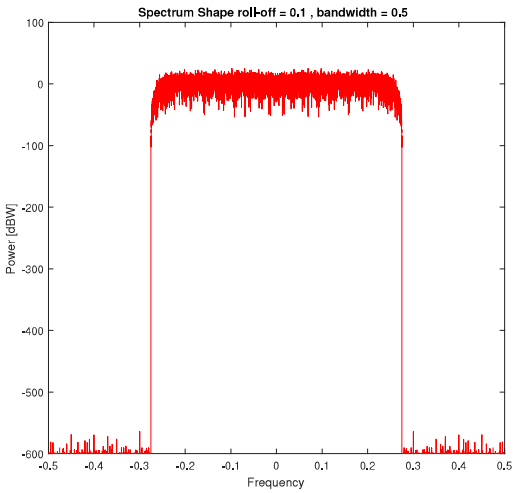


Fig 4.2: Spectrum shape of the input signal.

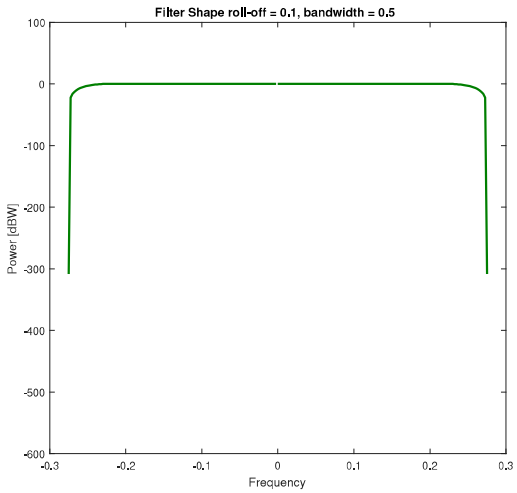


Fig 4.3: Filter shape.

In figure 4.2, there is the spectrum shape estimated by the penalized least squares approach in the setup simulation case. In the figure 4.3, the penalized least squares approach

will use the filter shape applied in the transmission, without any spectrum estimation.

Let's study the behavior in terms of mean squared error MSE measurements depending on the bandwidth, roll - off, generalized cross validation (section 3.5.2) and filter shape or spectrum operation mode.

The setup of the simulation consists in:
- number of samples: **1e5 sampes**
- Voltera memory = [**15, 3, 1**]
- generalized cross validation (gcv): **OFF**
- filter shape: **ON**
    - filter shape = **'raisedcosine'**
    - filter order = **3**
- spectrum shape: **OFF**
- error parameter: **Mean Squared Error (MSE)**
- variables: **signal bandwidth (GHz) and roll-off**

Mean squared error is defined as:

$$MSE = (1/n)\sum(Y - \hat{Y})^2$$

Where $\hat{Y}$ corresponds to the predicted values with the kernel estimation and **Y** corresponds to the measured and n corresponds to the number of samples.

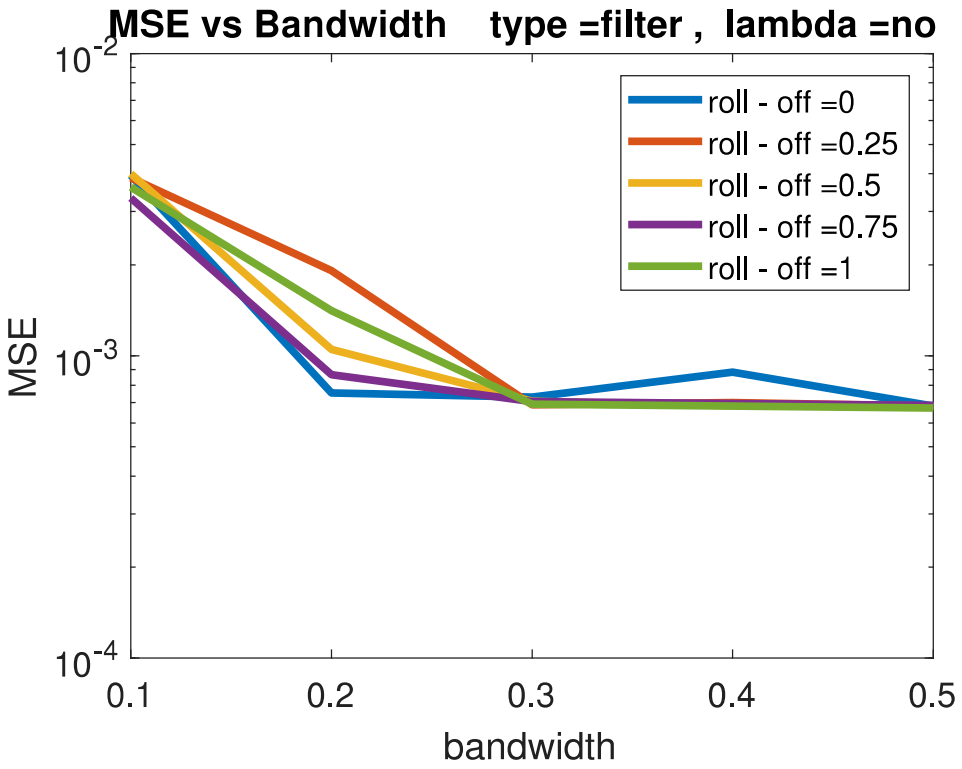


Fig 4.4: Mean squared error (MSE) vs singal bandwidth for different values of roll - off without using generalized cross validation and using the filter shape for the transfer function estimation in the penalized least squares approach.

As a general trend the MSE decreases with the rise of the signal bandwidth. The higher the signal bandwidth, the higher the level of information about the input signal which leads to better regularization functions. Then, the graph does not show significant correlation of the MSE with respect to the roll - off factor.

The setup of the simulation consists in:
- number of samples: **1e5 samples**
- Voltera memory = [**15, 3, 1**]
- generalized cross validation (gcv): **ON**
- filter shape: **ON**
    - filter shape = **'raisedcosine'**
    - filter order = **3**
- spectrum shape: **OFF**
- error parameter: **Mean Squared Error (MSE)**
- variables: **signal bandwidth (GHz) and roll-off**

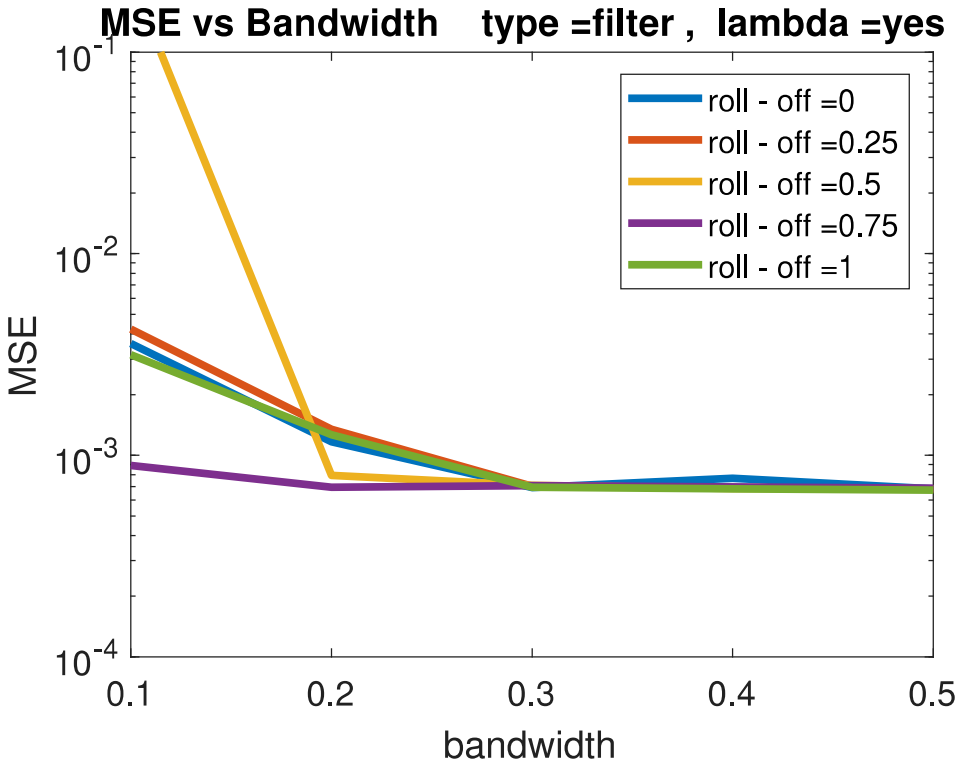


Fig 4.5: Mean squared error (MSE) vs signal bandwidth for different values of roll - off using generalized cross validation and using the filter shape for the transfer function estimation in the penalized least squares approach.

It can be appreciated in the case of high values of bandwidth, the mean squared error MSE tends to settled around a constant MSE error parameter. Even though there is not a significant difference between the use of generalized cross validation or not in the penalized least squares for the filter shape mode.

The setup of the simulation consists in:
- number of samples: **1e5 samples**
- Voltera memory = [**15, 3, 1**]
- generalized cross validation (gcv): **OFF**
- filter shape: **OFF**
    - filter shape = **'raisedcosine'**
    - filter order = **3**
- spectrum shape: **ON**
- error parameter: **Mean Squared Error (MSE)**
- variables: **sing bandwidth (GHz) and roll-off**

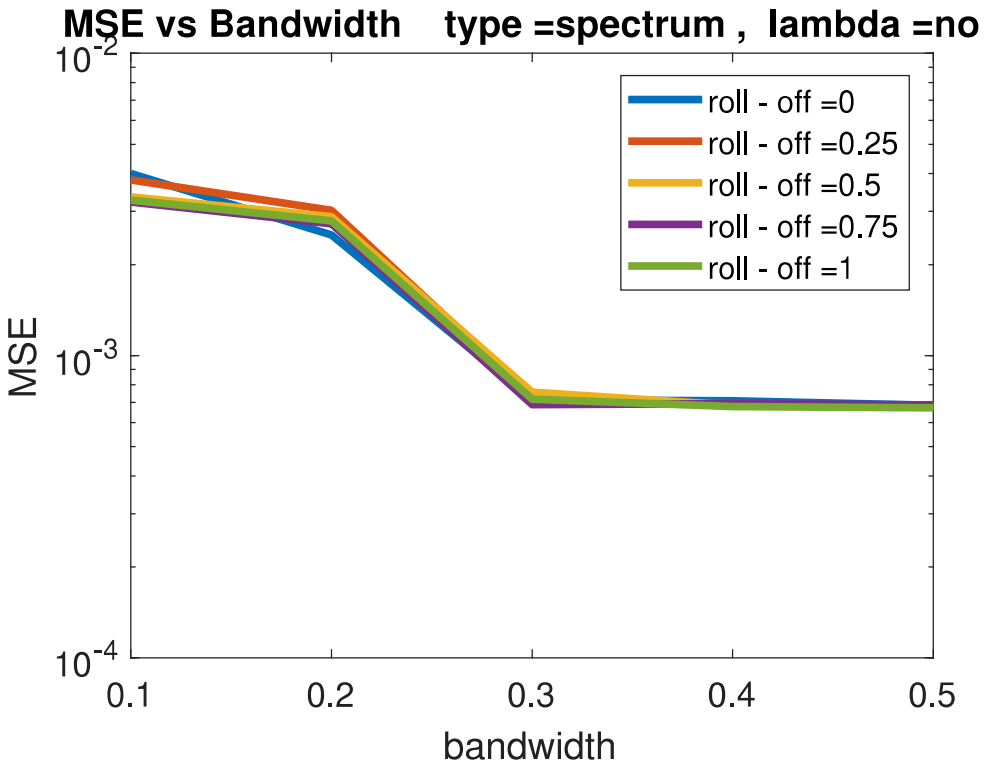


Fig 4.6: Mean squared error (MSE) vs signal bandwidth for different values of roll - off without using generalized cross vadilation and using the spectrum for the transfer function estimation in the penalized least squares approach.

As a general tendency the MSE decreases with the rise of the bandwidth.

The setup of the simulation consists in:
- number of samples: **1e5**
- kernels memory = [**15, 3, 1**]
- generalized cross validation: **ON**
- filter shape: **OFF**
    - filter shape = **'raisedcosine'**
    - filter order = **3**
- spectrum: **ON**
- error parameter: **Mean Squared Error (MSE)**
- variables: **bandwidth (GHz) and roll-off**

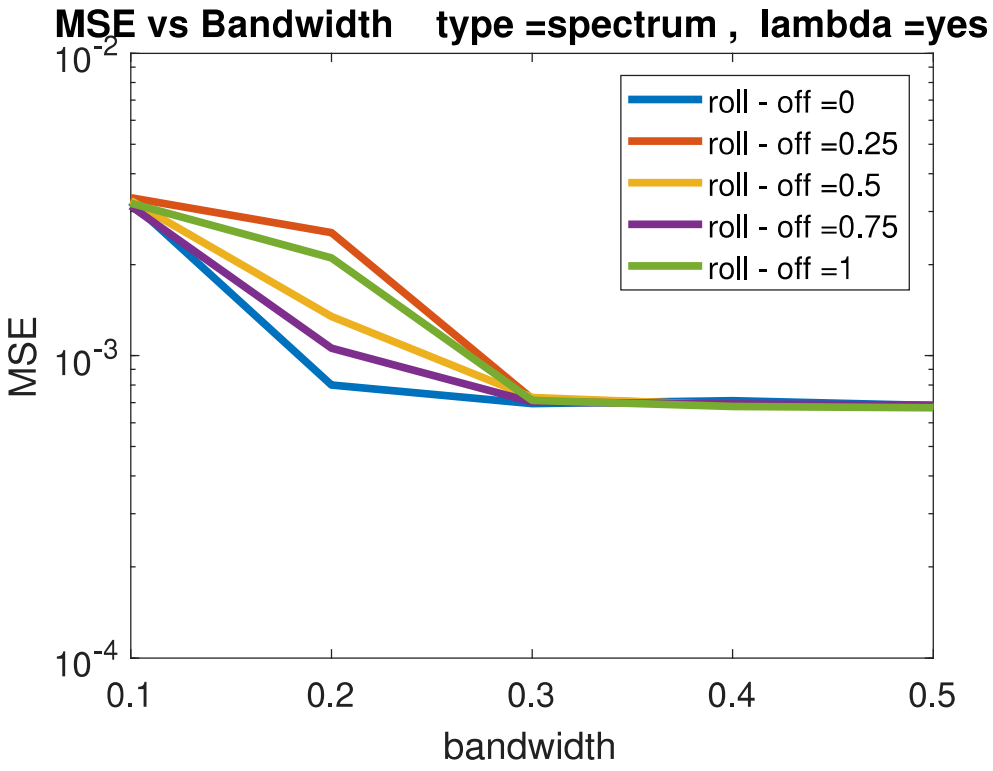


Fig 4.7: Mean squared error (MSE) vs bandwidth for different values of roll - off using generalized cross vadilation and using the spectrum for the transfer function estimation.

As a conclusion, the MSE decreases with the rise of the signal bandwidth. This fact is due to the nature of the regularization function, the higher the signal bandwidth the higher the signal information.

4.3 Penalized Least Squares Simulation

In order to study the pre-distorrion performance of the penalized least squares approach against the ordinary least squares approach, a simulation case is set up. The study case proposes an arbitrary nonlinear system, a negative exponential channel to study the estimation performance.

The setup of the simulation consists in:
- number of samples: **1e5 samples**
- Voltera memory = [**15, 3, 1**]
- filter shape: **ON**
    - filter shape = **'raisedcosine'**
    - filter order = **3**
    - filter roll - off = **0.1**
    - filter bandwidth = **0.5 GHz**
- variables: **signal to noise ratio SNR**

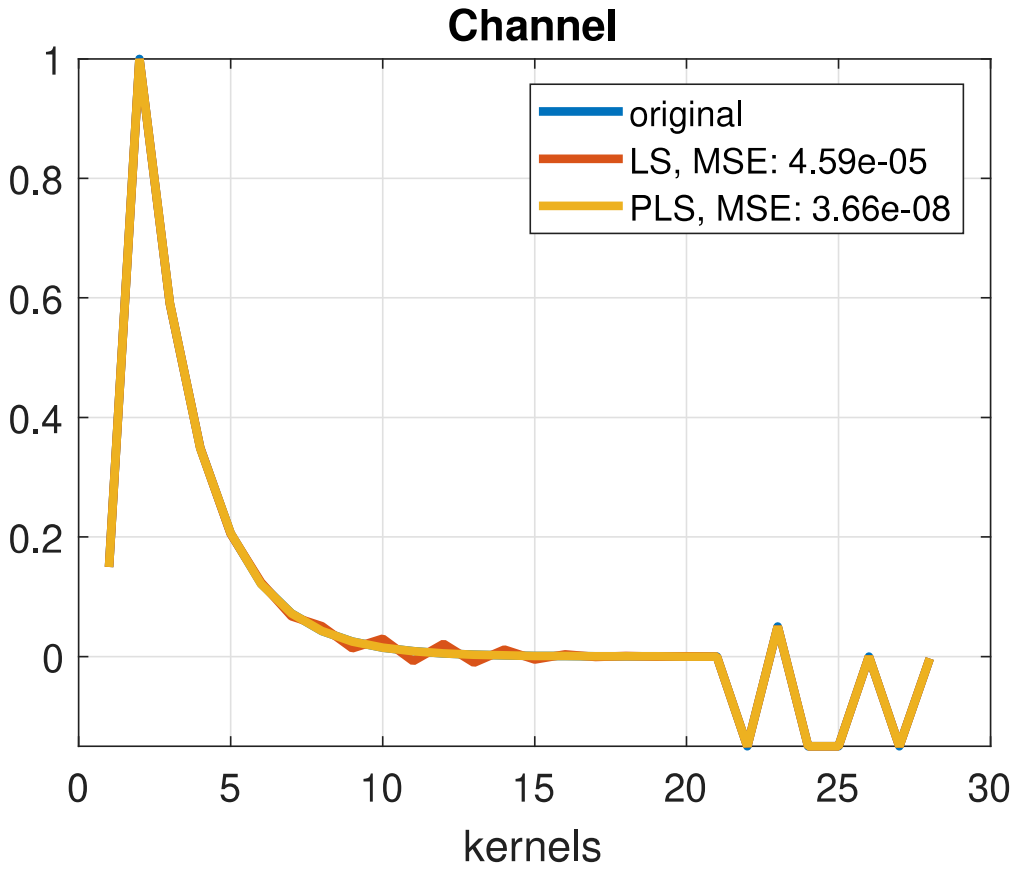


Fig 4.8: Representation of the original channel, LS estimation and PLS estimation for a **SNR = 30 dB** setup.

The previous figure 4.8 displays the different values of the Kernel's coefficients, for a system of this memory = [15, 3 ,1]. It can be appreciated that under good SNR regimes (SNR = 30 dB) the penalized least squares estimation (PLS) gives us a better estimation (MSE_{PLS} = **3.66e-08**) than the ordinary least squares estimation (OLS) (MSE_{OLS} = **4.59e-05**). Even though both estimations are good enough to be able to follow the original shape of the standard channel.

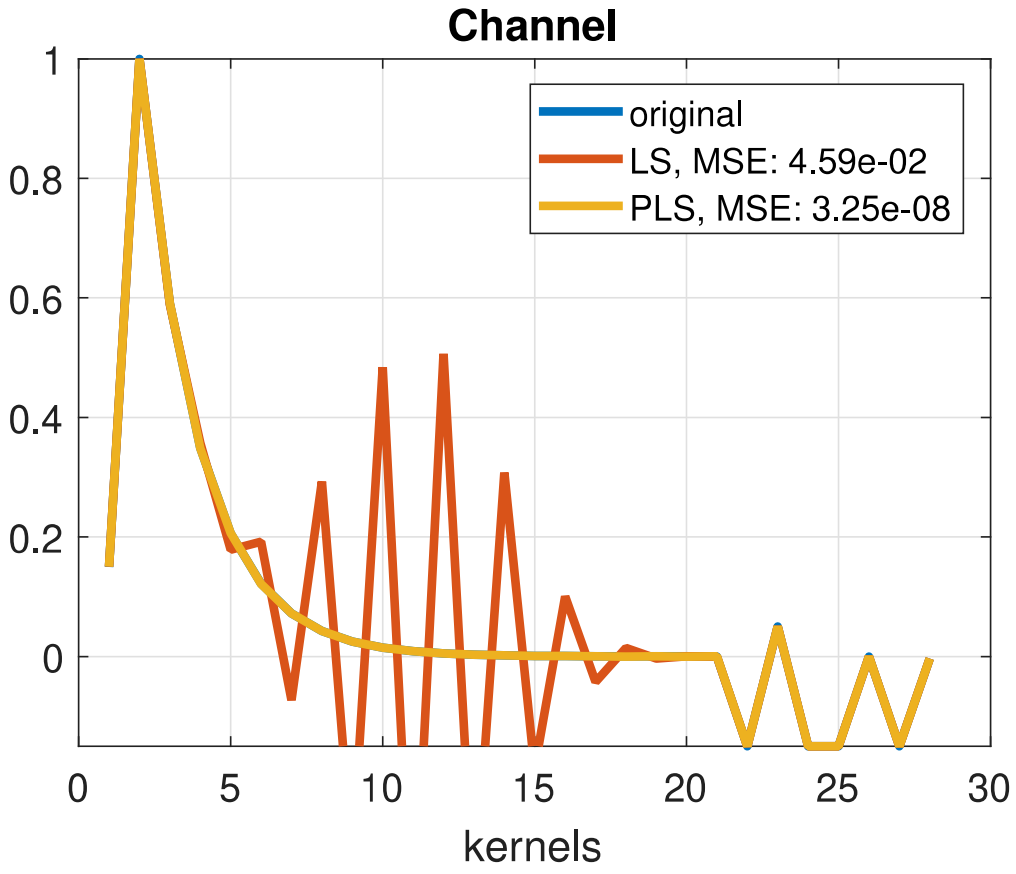


Fig 4.9: Representation of the original channel, LS estimation and PLS estimation for a **SNR = 15 dB** setup.

The previous figure 4.9 shows the different values of the kernels, for a system of this memory = [15, 3 ,1]. It can be noticed that under bad SNR regimes (SNR = 15 dB) the penalized least squares estimation (PLS) gives us a better estimation (MSE_{PLS} = **3.25e-08**) than the ordinary least squares (OLS) estimation (MSE_{OLS} = **4.59e-02**). In this case, the visual representation of the ordinary least squared is missing the track of the standard channel's shape while the penalized least squares keeps track of the standard channel.

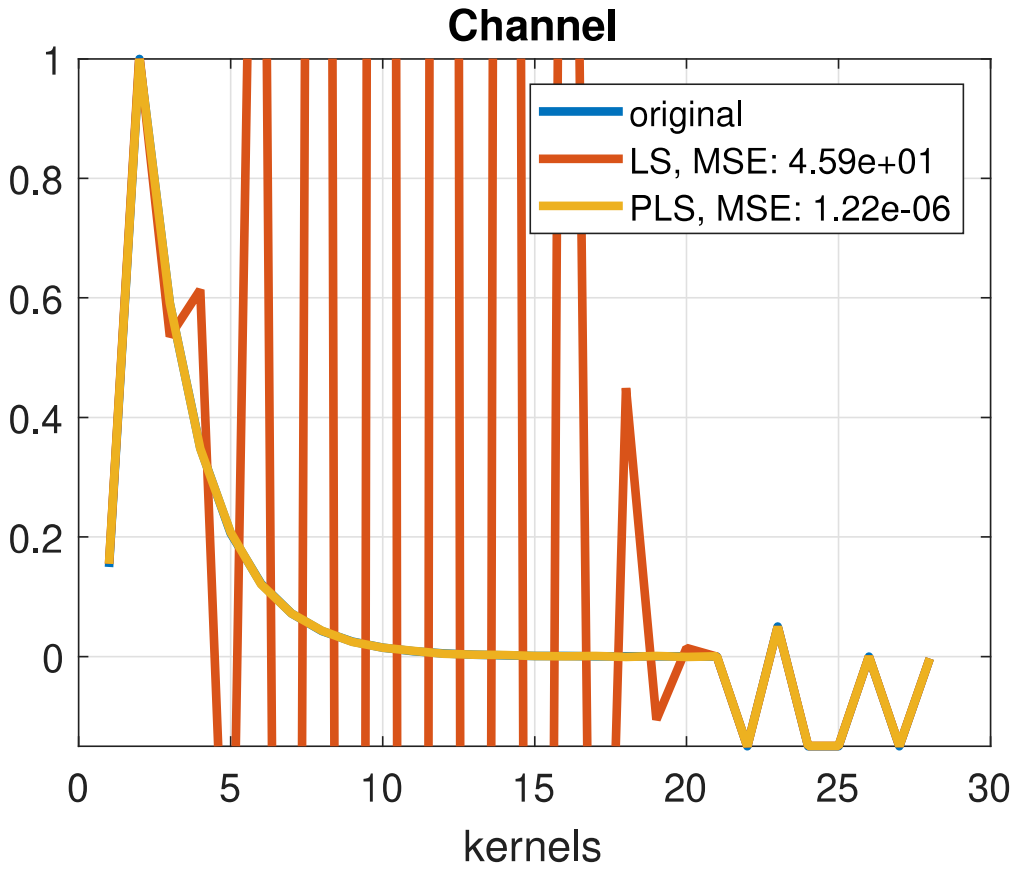


Fig 4.10: Representation of the original channel, LS estimation and PLS estimation for a **SNR = 0 dB** setup.

As the signal to noise parameter diminishes, the performance of the ordinary least squares (OLS) deteriorates while the performance of the penalized least squares (PLS) still keeps track of the standard channel.

The setup of the following simulation consists on:

- number of samples: **1e5 samples**
- Voltera memory: [**15, 3, 1**]
- signal to noise ratio range: **- 30dB to 50 dB**
- filter shape: **ON**
    - filter order = **3**
    - filter bandwidth: **0.2 GHz**
    - filter order: **3**
    - filter roll-off: **0.1**
    - filter shape: **'Gauss'**

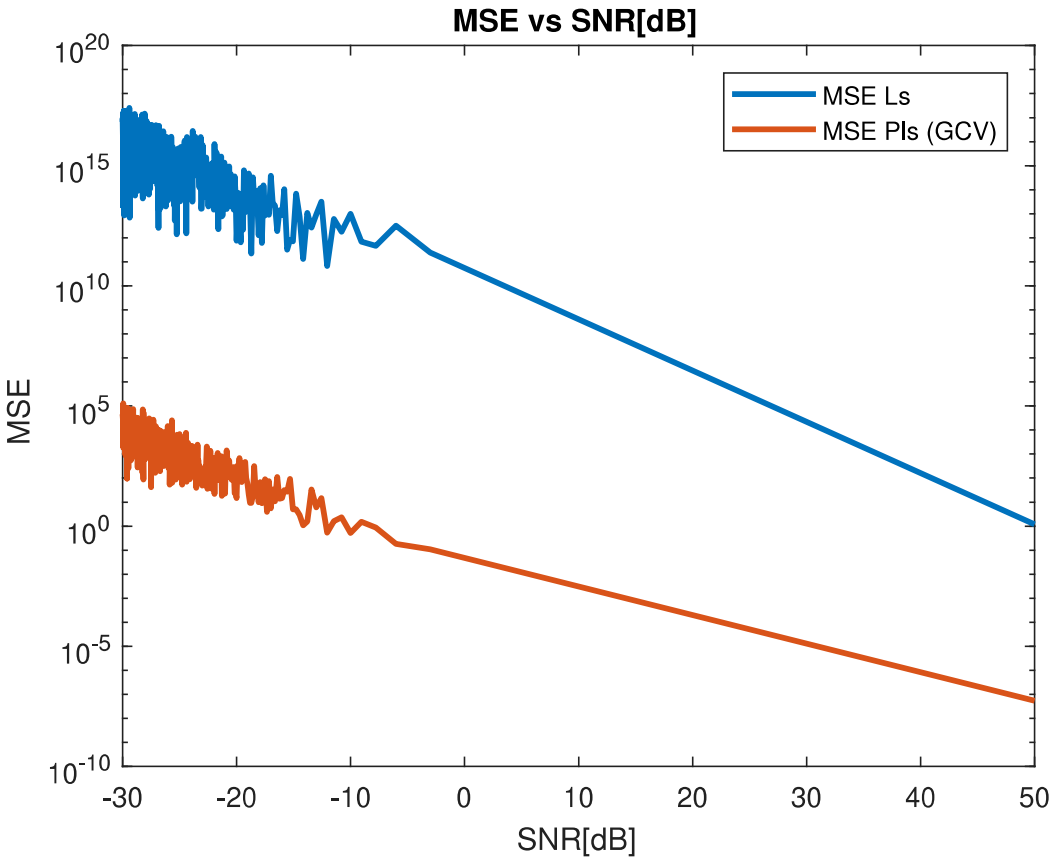


Fig 4.11: Mean squared error (MSE) error parameter for LS estimation and PLS estimation in a -30 dB to 50 dB range of SNR.

It can be recognized that the penalized least squares shows advantage against the least squares in both environments, in high SNR regimes and low SNR regimes.

4.4 Digital Signal Processing (DSP)

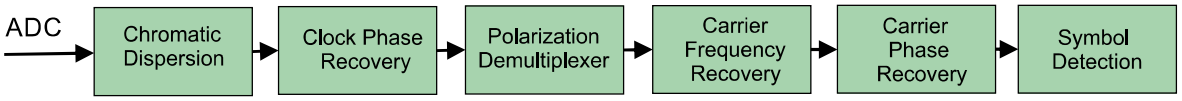


Fig 4.12: Flow chart of the digital signal processing (DSP).

This is the flow chart of a typical coherent DSP, the modules in figure 4.12 that are marked in green correspond to the modified modules object of this thesis. The rest of the modules are provided by the HHI library.

Before the flow chart, the first step is to sample and digitalize the received analog signals from the photodiodes of the coherent receptor (ADC). Then, it will be a chromatic dispersion (CD) compensation module (not required in the simulation setup). Then, the next problem the digital signal processing (DSP) faces is that the sampling rate clock at the receiver is asynchronous to the transmitter clock. The system is not sampling at the optimum pulse height, and the two clocks are not phase locked.

4.4.1 Clock Phase Recovery

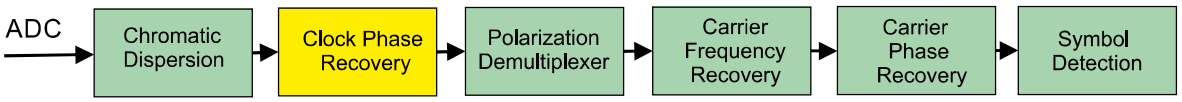


Fig 4.13: Clock Phase Recovery in Digital Signal Processing scheme.

After chromatic dispersion (CD) compensation,since there is not adaptiver equalizer in the setup to avoid removing channel effects that are useful in the pre-distortion process. It is necessary to acquire the clock phase of the data stream. The clock recovery algorithm finds the symbol clock frequency, selects the optimum sampling point and resamples the data accordingly (2 sa/s).

4.4.2 Polarization Demultiplexer

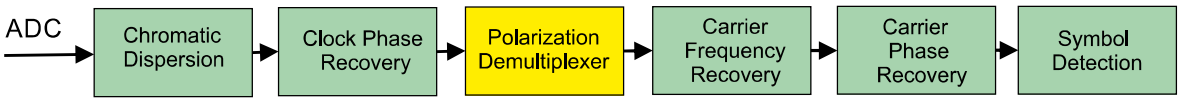


Fig 4.14: Polarization Demultiplexer in Digital Signal Processing scheme.

The purpose of this module is to implement a polarization demultiplexer; the main goal is not to alter the signal, for this reason there is not an adaptive equalizer in the dsp. In order to be able to pre-distort the signal, it cannot be equalized anyhow before; otherwise, the pre-distortion estimation will not work because all the effects of the channel would have been removed from the signal.

A performance test has been applied. The test consists in applying several polarization rotations to the signal, then check if the polarization demultiplexer is able to track and correct the rotation phases.

4.4.2.1 Cross - correlation Polarization Demultiplexer

After acquiring the right symbol clock, there are two signals (one for each polarization), but with mixed data due to the rotated state of polarization (SOP). The two polarization modes exchange energy through the mode coupling process during propagation, caused by the external perturbation due to bending or temperature variations.

The demultiplexer was performed such that the expected cross-correlation of the two polarizations is minimized [16]. The polarization rotation is estimated by determining the minimum value of the cross-correlation between the two complex inputs to the demultiplexer (E_x, E_y) as a function of a polarization rotation phase [17] [18].

This can be calculated in parallel with a finite number of spaced values of θ on the interval $-\pi/4 < \theta \leq \pi/4$, since:

$$< E_x(\theta)E_y(\theta)^* >= cos^2(\theta) < E_xE_y^* > -sin^2(\theta) < E_x^*E_y > +sin(\theta)cos(\theta) < |E_y|^2 - |E_x|^2 > \tag{4.1}$$

Then the optimal value of θ_{opt} is determined as the angle that minimizes the previous expression. Then, the Stokes polarization rotation is described as follows:

$$R = \begin{bmatrix} cos(\theta_{opt}) & sin(\theta_{opt}) \\ -sin(\theta_{opt}) & cos(\theta_{opt}) \end{bmatrix}$$

Important results of the cross - correlation polarization demultiplexer can be appreciated in the comparative graphs 4.15 and 4.16 that study as well as the penalty error vector magnitude (EVM) to be applied in relationship with the signal-to-noise ratio (SNR).

The error vector magnitude (EVM) describes the vector errors between the received constellation and the ideal constellation. In this case, the reference vectors are those that join the origin with the ideal positions of the constellation points and the error vector, which are those vectors from the ideal position of the constellation diagram to the points of the received constellation.

$$EVM(\%) = \sqrt{\frac{A_{error}}{A_{reference}}} x100$$

Where A_{error} is the module of the error vector and $A_{reference}$ is the module of the highest power point in the reference signal constellation.

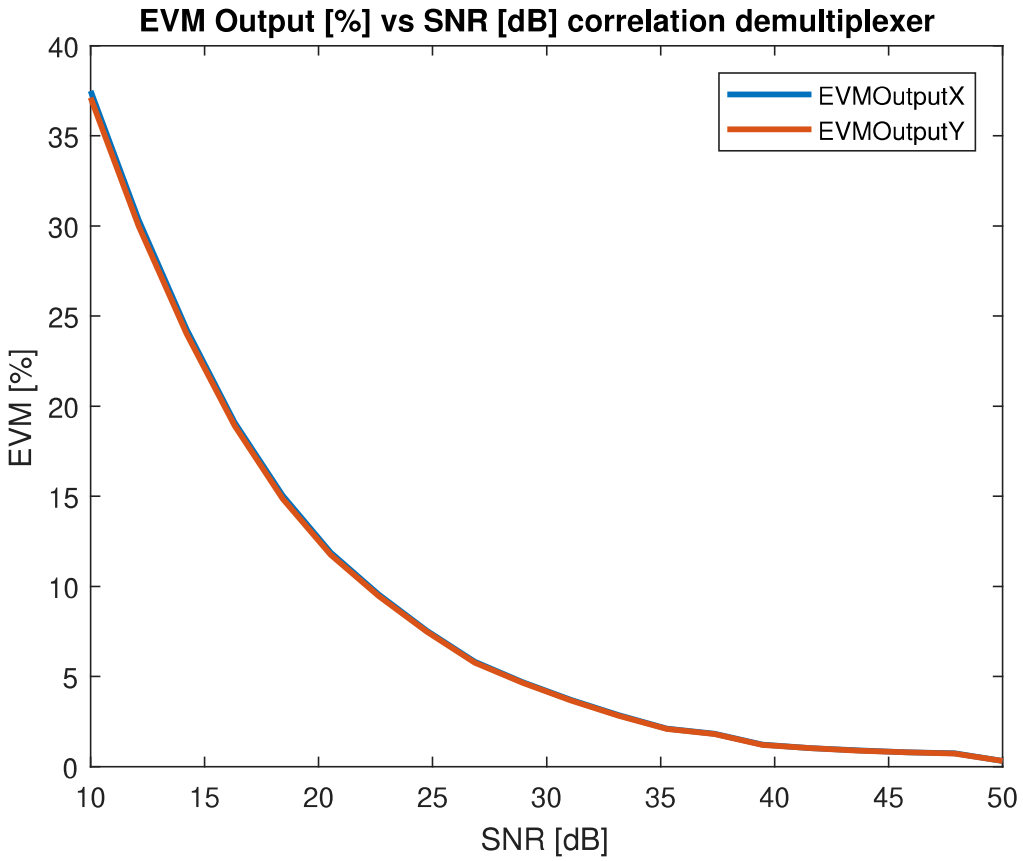


Fig 4.15: The output Error Vector Magnitude (EVM) of both polarization **X** and **Y** in relationship with SNR for the cross - correlation polarization demultiplexing.

There is a negative trend in the EVM Error Vector Magnitude as the SNR increases.

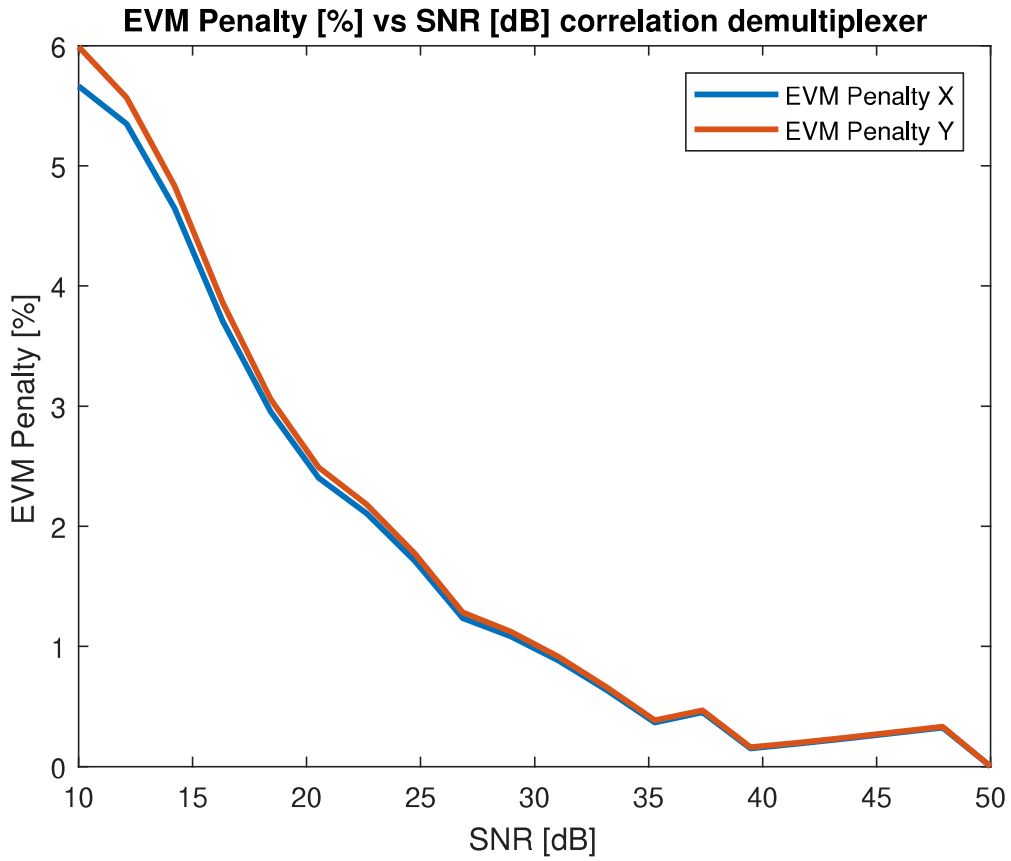


Fig 4.16: Error Vector Magnitude (EVM) penalty of both polarizations **X** and **Y** in relationship with SNR for the cross - correlation polarization demultiplexing.

The EVM penalty has been computed as the difference between the EVM of the output and the EVM of the input.

$$EVM_{PENALTY}(\%) = EVM_{output}(\%) - EVM_{input}(\%)$$

There is a high EVM penalty for low values of SNR. Then, as the SNR increases the EVM penalty tends to decline.

4.4.2.2 State of Polarization (SOP) demultiplexer

This section presents a technique for polarization demultiplexing of arbitrary complex-modulated signals. The technique uses the data in Stokes space [19][20].

Stokes parameters

In order to apply the necessary correction in terms of polarization rotation it is interesting to work with Jones parameters. In order to translate Stokes vector into a Jones vector it is critical to understand that the fourth - dimensional Stokes vector can be projected onto a three - dimensional vector leaving out the intensity (S_0). However, the Jones vector is still constituted of two - dimensional parameter.

The computation of the Stokes parameters S_0, S_1, S_2, S_3 from the signal field E_X in the **X** polarization and the E_Y in the **Y** polarization is represented in the following equations:

$$S_0 = E_X E_X^* + E_Y E_Y^*$$

$$S_1 = E_X E_X^* - E_Y E_Y^*$$

$$S_2 = E_X E_Y^* + E_X^* E_Y$$

$$S_3 = jE_X E_Y^* - jE_X^* E_Y$$

The relationship of the Stokes parameters S_0, S_1, S_2, S_3 to intensity **I** and polarization ellipse coordinates in shown below.

$$S_0 = I$$

$$S_1 = Ipcos(2\theta)cos(2\epsilon)$$

$$S_2 = Ipsin(2\theta)cos(2\epsilon)$$

$$S_3 = Ipsin(2\epsilon)$$

Where **I**, **p**, 2θ and 2ϵ are fourth - dimensional coordinates. Then, leaving out of the system the parameter **I** which stands for the intensity, a three - dimensional coordinates are left. Finally, applying the following transformations it can be obtaind the two - dimensional coordinates of the Stokes vector.

$$2\theta = atan\frac{S_2}{S_1} \qquad 2\epsilon = atan\frac{S_3}{\sqrt{S_1^2 + S_2^2}}$$

Lens - like objects

The lens-like object defines an ellipsoid which normal vector contains the average of the polarization states of transmission [19]. This ellipsoid is defined by three axis, in this case, the shortest axis corresponds to a normal vector pointing to the polarization states of transmission . This ellipsoid and its normal vector can be found from least squares fit of the data including symbols and transitions.

Poincaré Sphere

The Poincaré sphere, which was proposed by Jules Henri Poincaré in 1891-1892 [21], is a favorable approach to represent polarized light. This approach is a technique based on coordinates: each point on the sphere corresponds to a certain state of polarization SOP. Because a state of polarization SOP is represented by a point in the sphere, a evolution of polarization can be tracked as a continuous path on the Poincaré sphere.

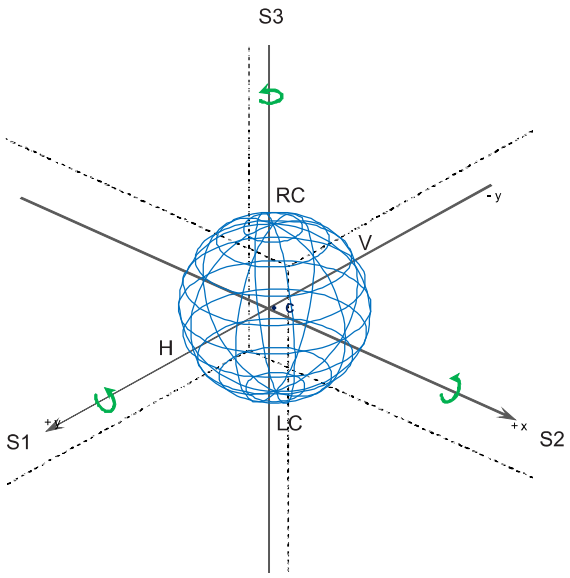


Fig 4.17: Poincaré Sphere

Where the intersection of the Poincaré sphere with $+S_3$ in figure 4.17 stands for a right circular polarization (RC) and the intersection with -S_3 in figure 4.17 denotes left circular polarization (LC).

Then, the intersection of the Poincaré sphere with $+S_1$ ($+y$ in figure 4.17) stands for horizontal polarization (H) and the intersection with -S_1 stands for vertical polarization (V). Finally, the intersection of the Poincaré sphere with $+S_2$ ($+x$ in figure 4.17) stands for a +45 ° linear polarization and the intersection with -S_2 stands for a -45 ° linear polarization.

Method

The idea is to rotate the lens-like object into the S_2S_3 plane so that the normal vector is aligned with the S_1 axis ($+x$ in figure 4.17) of the Poincaré sphere, which represents the horizontal and vertical polarization, which are, ideally, the initial polarizations for the **X** and **Y** components.

In order to rotate the ellipsoid where the states of polarization lie into the plane S_2S_3. The algorithm should apply the sequential rotation, first around the S_3 axis of the Poincaré sphere. Then, around the S_2 axis of the Poincaré to finally lie the states of polarization ellipsoid into the S_2S_3 plane. (the rotation algorithm is the contribution of this thesis to the SOP demultilpexer).

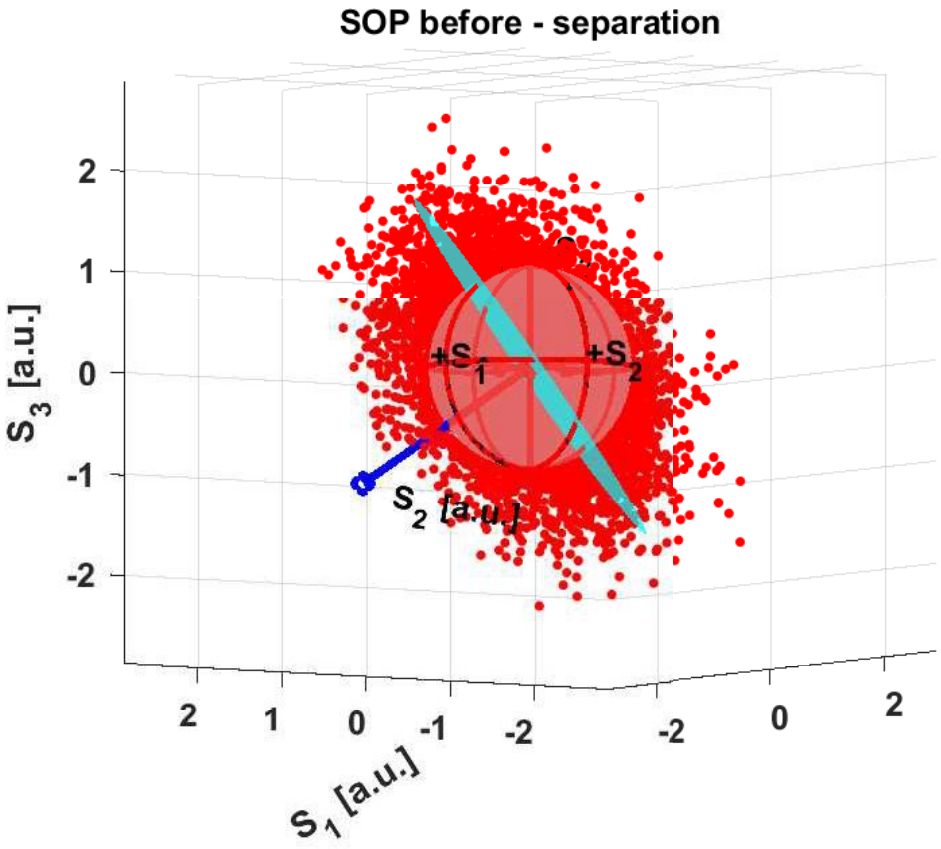


Fig 4.18: state of polarization (SOP) before rotation.

It can be appreciated that all the different states of polarization lying in a ellipsoid or lens-like structure in the figure 4.18. The ideal case, is when the normal vector of the plane is aligned with the S_1 ($+x$ in figure 4.17) axis of the Poincaré sphere, this is, the plane should be lying in the S_2S_3 plane. Otherwise, the algorithm has to correct the vector position.

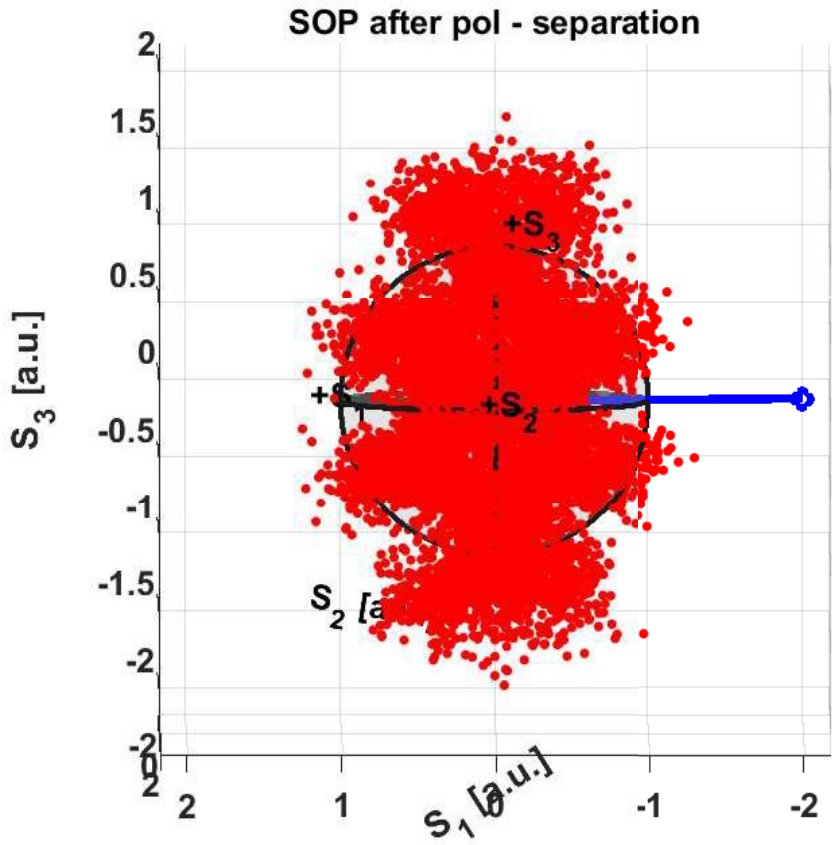


Fig 4.19: state of polarization (SOP) after rotation.

Finally, the two polarizations are demultiplexed into a vertical and horizontal polarizations.

The following figures 4.20 and 4.21 display important results from the polarization demultiplexing in Stokes space. The results consist on a performance test in terms of error vector magnitude (EVM) as well as the penalty EVM in relationship to the signal-to-noise ratio SNR as mentioned in section 4.4.2.1.

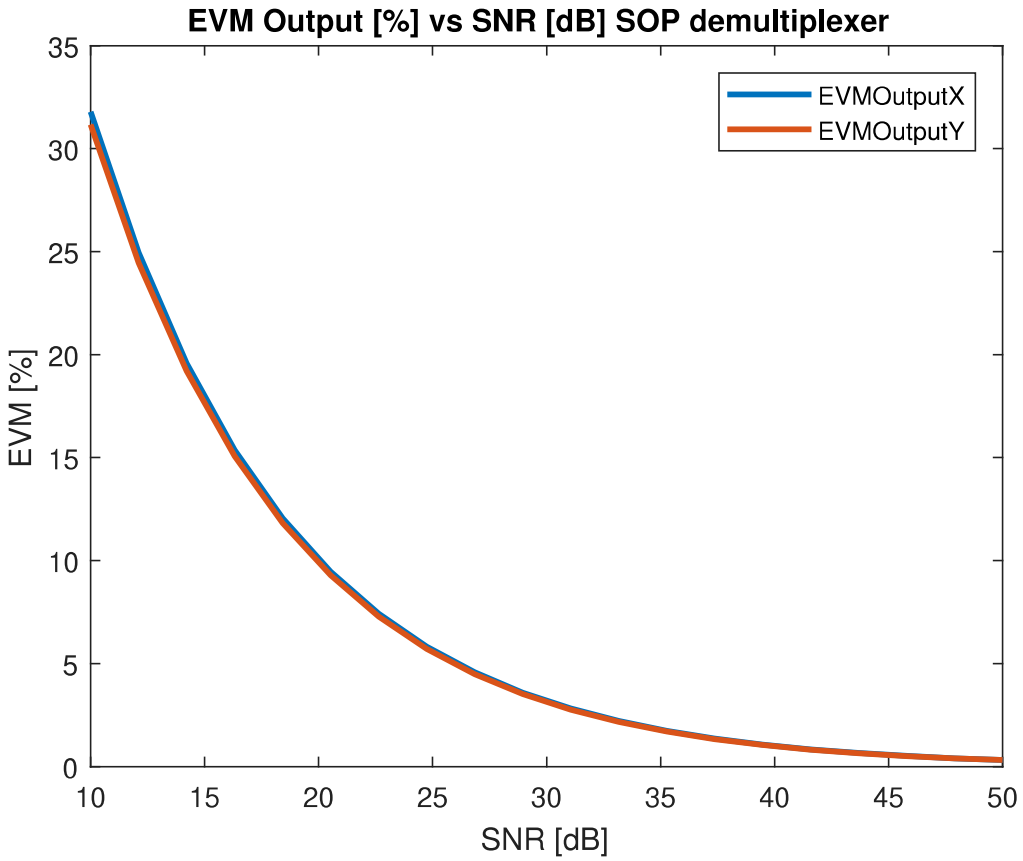


Fig 4.20: the output Error Vector Magnitude (EVM) of both polarization **X** and **Y** in relationship with SNR for the polarization demultiplexing in Stokes space.

There is a negative trend in the Error Vector Magnitude EVM as the signal-to-noise SNR rises.

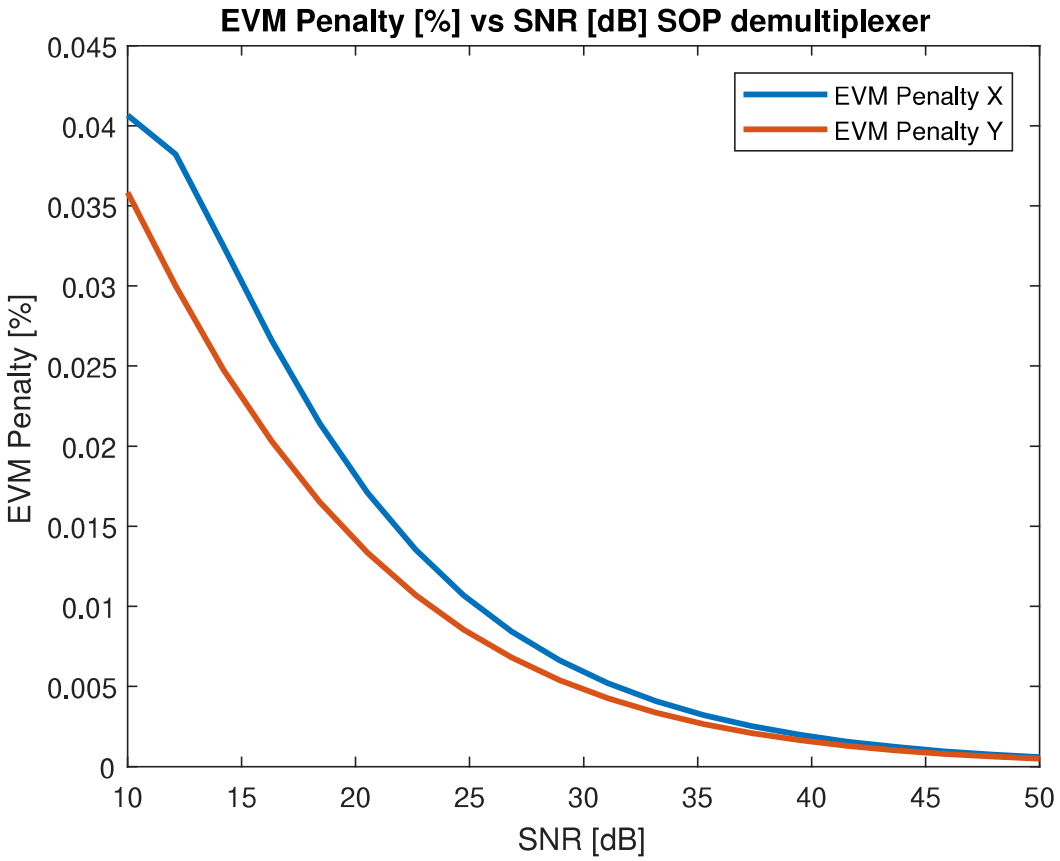
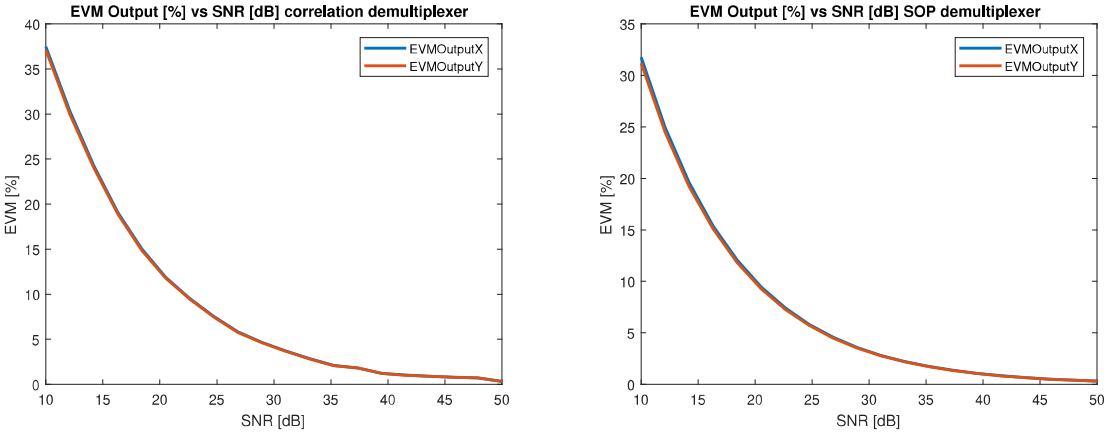


Fig 4.21: Error Vector Magnitude (EVM) penalty of both polarizations **X** and **Y** in relationship with SNR for the polarization demultiplexing in Stokes space.

$$EVM_{PENALTY}(\%) = EVM_{output}(\%) - EVM_{input}(\%)$$

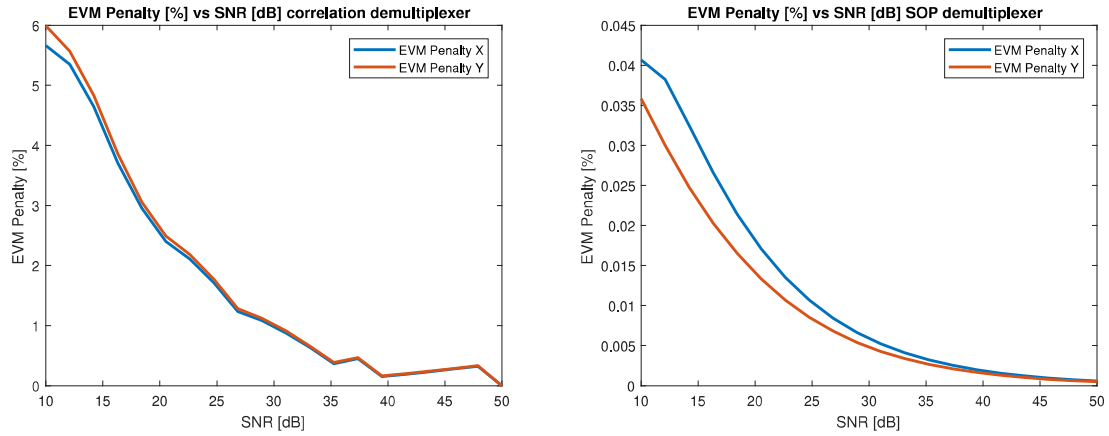
Even though the overall EVM penalty is very smallki, there is a higher EVM penalty for low values of SNR. Then, as the SNR increases, the EVM penalty tends to decrease.

4.4.2.3 Performance Comparison



(a) (b)

Fig 4.22: Output Error Vector Magnitude (EVM) comparison between the cross - correlation (a) and the state of polarization SOP (b) demultiplexing approach.



(a) (b)

Fig 4.23: Error Vector Magnitude (EVM) penalty comparison between the cross - correlation (a) and the state of polarization SOP (b) demultiplexing approach.

The EVM penalty has been computed as the difference between the EVM of the output signal and the EVM of the input signal.

In comparison with the previous cross - correlation approach, the EVM penalty percentage is significantly lower in the state of polarization SOP approach. For this reason, the state of polarization SOP demultiplexing method is applied for the different simulations and experiments in this thesis.

4.4.3 Carrier Frequency Recovery

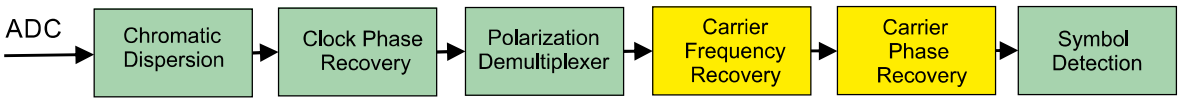


Fig 4.24: Carrier Frequency Recovery Digital Signal Processing scheme

This module enables the removal of the frequency offset or residual frequency error due to misalignment between the transmitter laser and the local oscillator in the receiver, which effect is represented by the addition of phase noise.

4.4.4 Carrier Phase Recovery

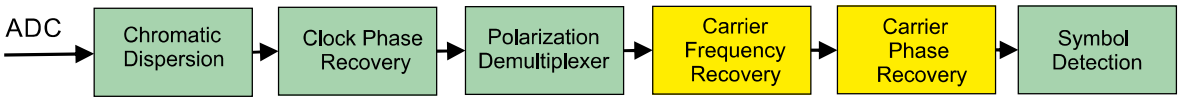


Fig 4.25: Carrier Phase Recovery in Digital Signal Processing scheme

This module analysis the performance of different carrier phase estimation algorithms such as blind phase search (averaging block based) or Viterbi sliding window (sliding window averaging) working at 2 sa/s. The Viterbi sliding window has been applied for this thesis. This is the scheme in the adaptation of the carrier phase recovery module.

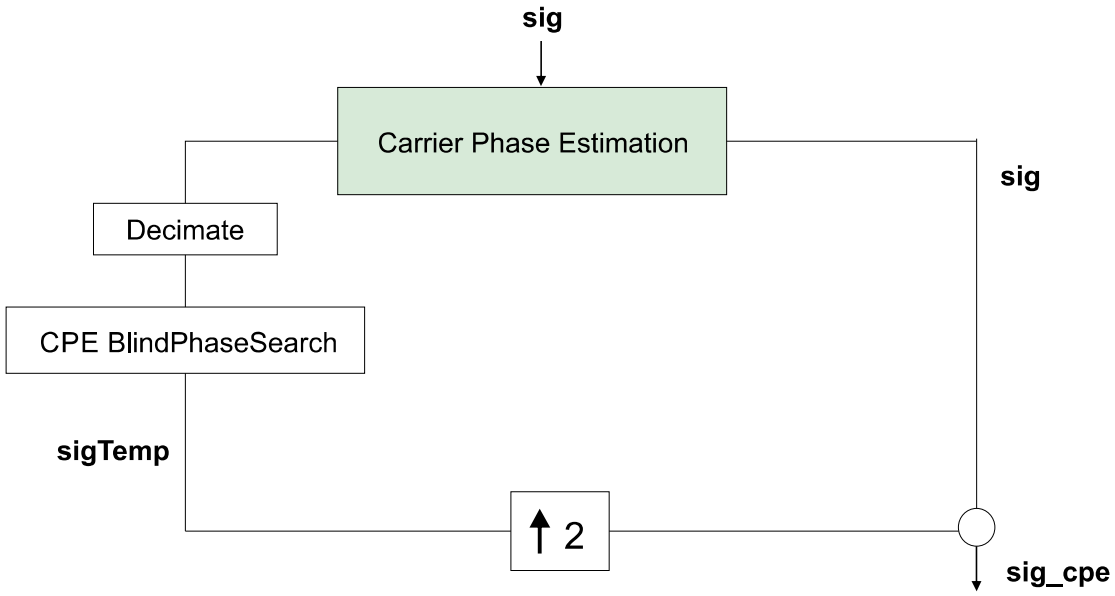


Fig 4.26: Schematic of carrier phase estimation working at 2sps

This is the scheme of the modified version of the carrier phase estimation module working at 2 sa/s. The module CPE BlindPhaseSearch is provided by the HHI library is a carrier phase estimation based on Viterbi sliding window (sliding window averaging).

Once the signal arrives to the Carrier Phase Estimation it is divided in two branches. On one hand, the signal is decimated to enter the CPE BlindPhaseSearch module that works at 1sa/s. On the other hand, the right branch in the figure 4.26 corresponds to the original signal without decimation (sig). The left branch of the figure 4.26 corresponds to the phase estimations upsampled (sigTemp). Finally, the two branches are multiplied and then the phase is corrected (sig cpe).

The performance of this module has been tested as a whole dsp in the experiment.

Chapter 5

Experiment

This chapter explains the experimental setup in the HHI laboratory. This scenario is used to test the performance of the regularization algorithm in penalized least squares approach for nonlinear system identification.

Experiment procedure

The experiment procedure is the following:

First, an input signal is transmitted to be received without using digital pre-distortion at the transmitter of the coherent system, the received signal is named as measured.

Then, the input signal, the received signal (measured) and the Volterra memory are used to estimate the Kernel coefficients of the channel for future pre-distortion.

Next, the same input signal will be pre-distorted by the previously calculated Kernel coefficients at the transmitter level. The received signal is named as simulated.

Finally, there is a comparison between the non pre-distorted received signal (measured) and the pre-distorted received signal (simulated).

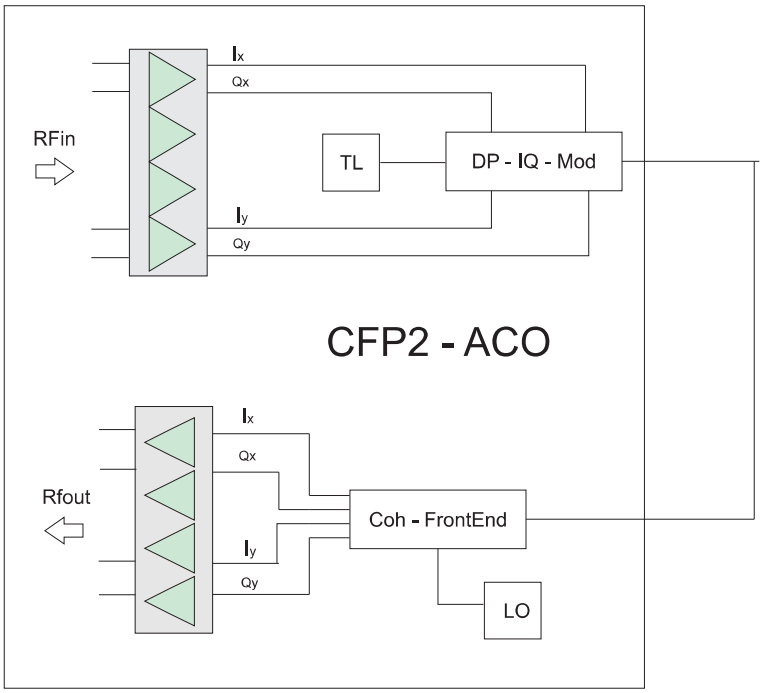


Fig 5.1: Schematic of the experiment setup.

The experiment consists in a coherent transmission system set by interconnecting the transmitter module and the receiver module of the transceiver CFP2 - ACO module of the HHI laboratory.

5.1 Experiment schematic setup

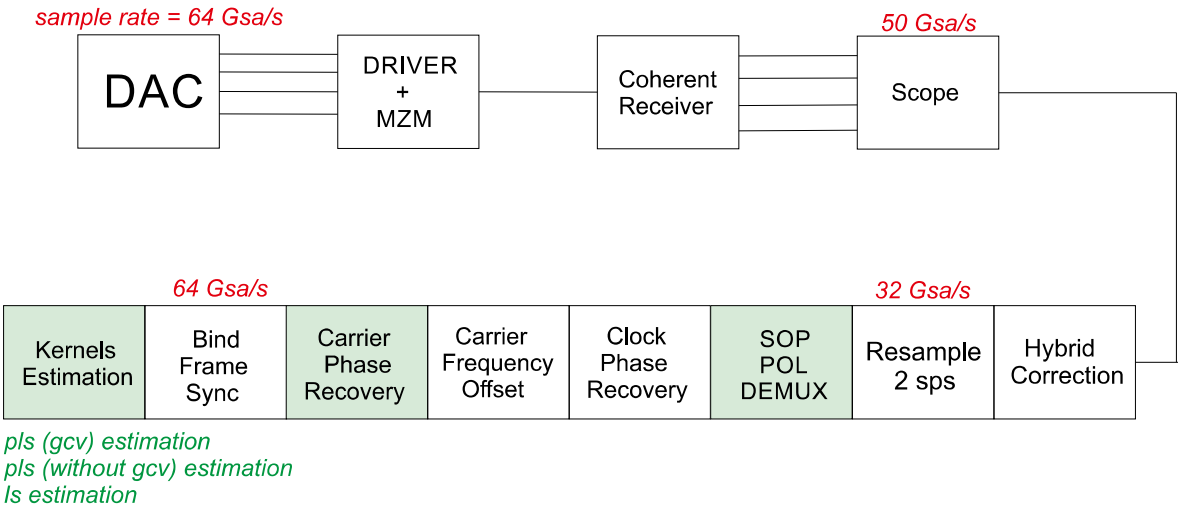


Fig 5.2: Complete scheme of the experimental system setup.

The described setup in figure 5.2 shows applied to test the performance of the penalized Kernel estimator. The figure 5.1 is included in the top branch in figure 5.2. The green marked modules in the bottom branch of figure 5.2 correspond to the modified modules for the purpose of this thesis, the rest of the modules and instrumentation are provided by the HHI library and HHI laboratory.

First, the system generates a **PDM - 16QAM** at **16 GBaud** symbol rate and **20 GHz** bandwidth signal. The generated signal is passed through a digital to analog converter (DAC), then each component of each polarization (I_x, Q_x, I_y, Q_y) is amplified by the driver amplifier to be modulated by a **DP - IQ - Mod**.

The acquired or measured signal by the oscilloscope (end of the top branch in the figure 5.2) is processed by the digital signal processing modified in this thesis; this digital signal processing is presented in the bottom branch of the figure 5.2. The first step of the digital signal processing consists on a hybrid correction module, which applies an orthogonalization procedure in order to compensate for a quadrature and in-phase imbalance in the optical coherent receiver, this module is provided by the HHI library. Then the signal is resampled to a sample rate of 32 Gsa/s (2 sa/s).

After that, the signal goes through a state of polarization (SOP) demultiplexer which has been modified in this thesis, followed by a clock phase recovery module from the HHI library. Once the signal has been demultiplexed, both polarizations go through a carrier frequency offset module of the HHI library; its main goal is to compensate the frequency offset. Then, the carrier phase estimation module has been updated to perform at two samples per symbol, it is applied to estimate and recover the phase by using the Viterbi sliding window.

At this point, the only thing missing is the synchronization between the measured signal and the reference signal in order to be able to perform an estimation of the kernel coefficients. Since the reference signal has a sample rate of 64 Gsa/s, the current digital signal process (bottom branch of figure 5.2) should upsample to 64 Gsa/s from the previous sample rate 32 Gsa/s.

Once at the Kernel estimator module (regularization function for the penalized least squares correspond to this thesis contribution), the Kernel estimator requires different parameters such as; the input signal, the output signal and memory of the different Volterra orders to be able to estimate the system. Here, there are the options of using ordinary least squares (ls) approach, penalized least squares (pls) approach or the penalized least squares (pls) with generalized cross validation (gcv).

5.2 Parameters setup of the experiment

The setup of the experiment consists in the following system parameters:

- modulation format: **PDM - 16QAM**
- pulse shape: **'root raised cosine'**
- roll-off factor: **0.1**
- laser wavelenght: **1550.1 nm / 193.4 THz**
- Volterra memory: [**200, 6, 6**]
- number of symbols: 2^{15} **symbols** (DAC)
- number of samples: 2^{17} **samples** (DAC)
- symbol rate: **16 GBaud**
- oscilloscope IDN: **TEKTRONIX, DPO72004, CF:91.1CT**
- channel bandwidth: **20.0 GHz**
- channel deskew: **0.00, -0.02, 0.05, -0.021 ns**

Where channel deskew stands for the compensation of the relative delay differences of the four paths, the largest delay in the paths is considered the reference. The cause of this delays is the different physical lengths of the cables (optical and electrical) from the receiver hybrid to the scope.

5.3 Digital Signal Processing (DSP)

This section corresponds to the modified digital signal processing shown in the bottom branch of the figure 5.2. First, the received signals from the photodiodes of the coherent receiver are sampled at 32 Gsa/s.

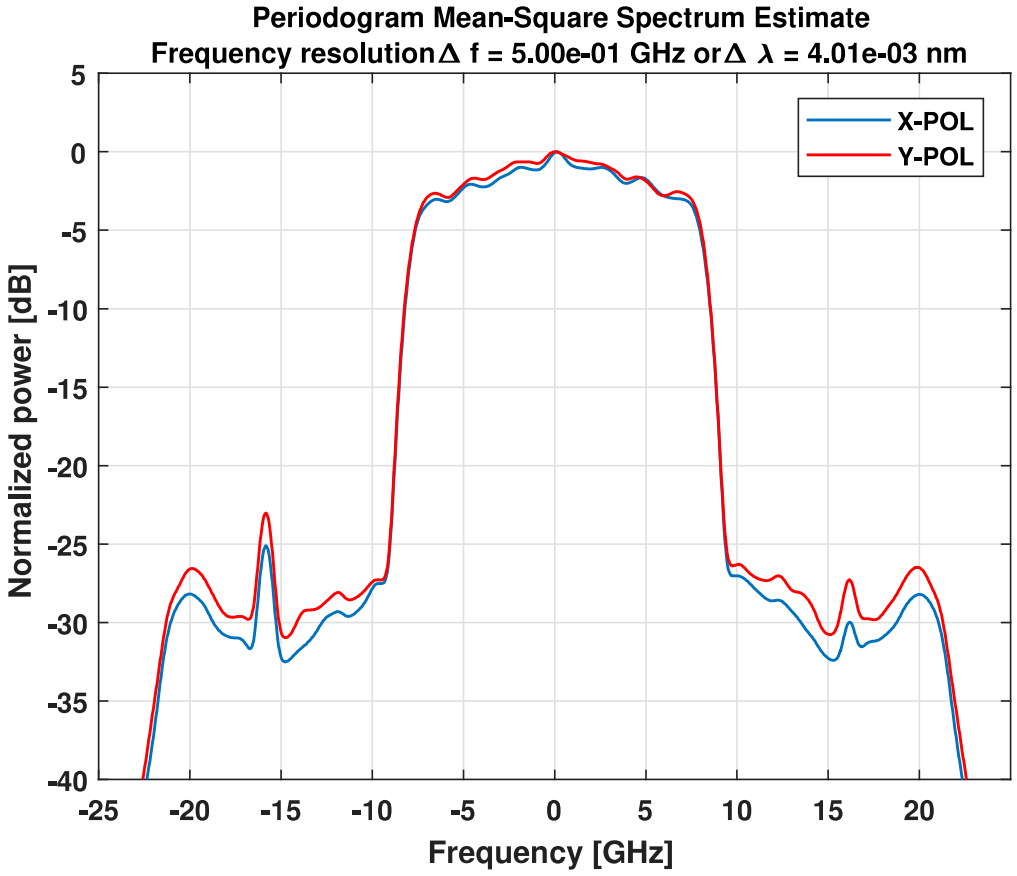


Fig 5.3: Spectrum of the **X** and **Y** component of the signal at the oscilloscope.

The measured signal spectrum corresponds to the signal acquired at the oscilloscope (at the end of the top branch shown in the figure 5.2), it can be recognized the root raised cosine shape with background noise. It can be observed that the signal power is attenuated in the surroundings of the band of interest as well as frequencies components beyond the channel bandwidth **20 GHz** due to the bandwidth of the oscilloscope.

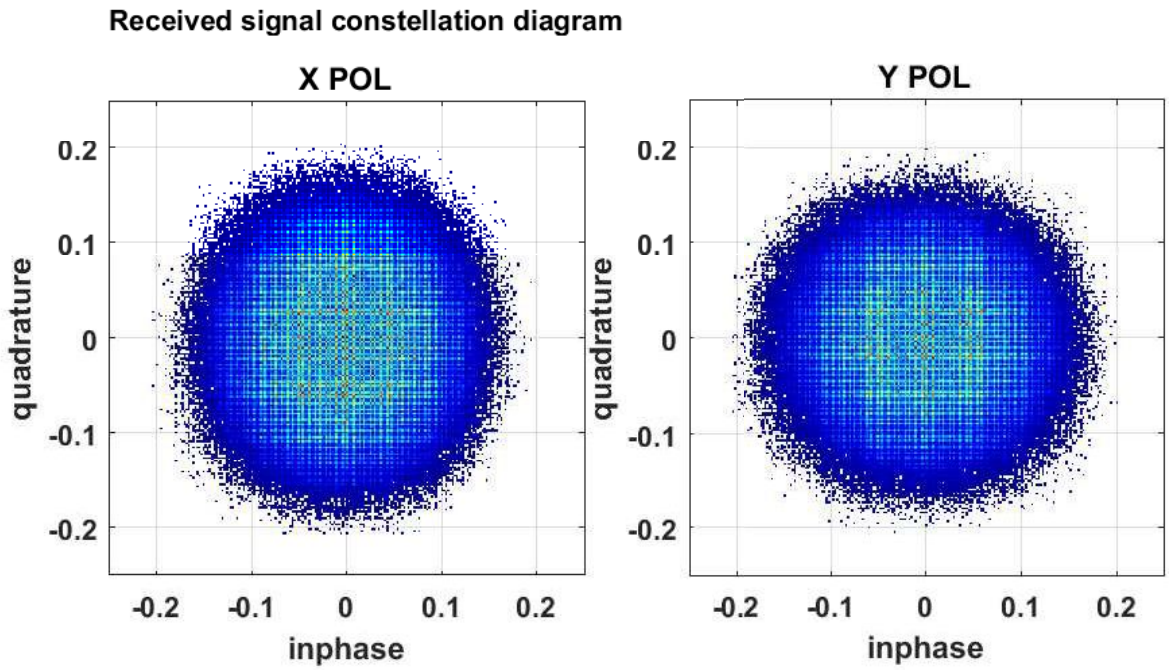


Fig 5.4: Constellation diagram of the received signal before dsp.

The received constellation at the input of the digital signal processing chain is shown in figure 5.4. Due to different path gains, the constellations are not fully circular but there is an IQ imbalance, specially in the case of the **X** polarization that has an ellipse shape.

5.3.1 Hybrid Correction

The hybrid correction module is intended to compensate the IQ imbalance mentioned before.

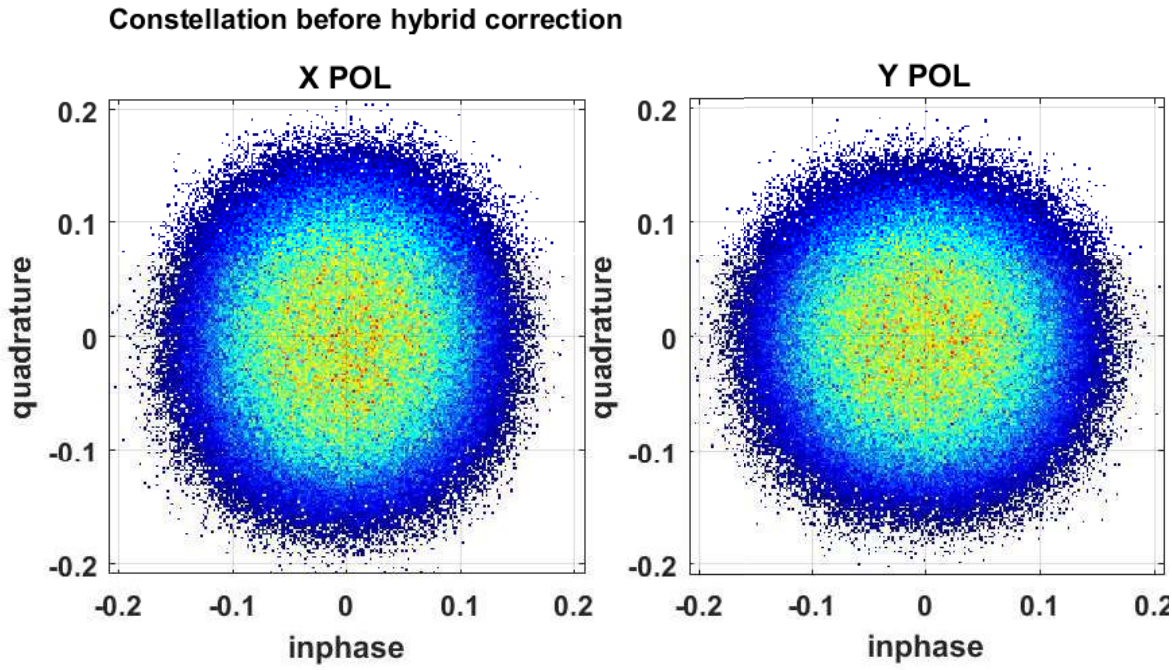


Fig 5.5: Constellation diagram before hybrid correction (2 sa/s).

After the hybrid correction module, the amplitudes of **I** and **Q** have been compensated, specially the case of the **X** polarization. Instead of being an ellipse, it has taken a circular or balanced shape.

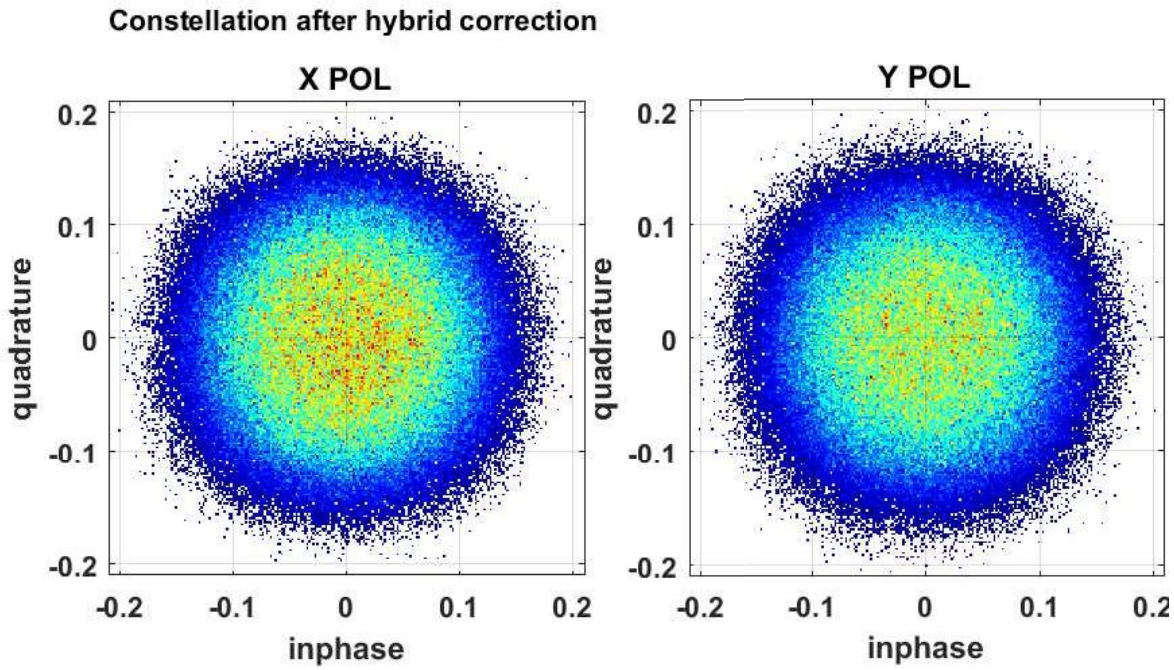


Fig 5.6: Constellation Diagram after hybrid correction (2sps).

5.3.2 Polarization Demultiplexer

One of the goals of this thesis is to finish a polarization demultiplexer module based on state of polarization which is property of the HHI institute. The main function of an equalizer is to reverse the effects of the channel upon the signal, in our case there is not adaptive equalizer.

The idea is to maintain the channel effects for the pre-distortion module to take those effects into account. In order to be able to pre-distort the signal, it cannot be equalized anyhow before, otherwise the pre-distortion estimation will not work.

5.3.2.1 Polarization Demultiplexer Results

The previous section 4.4.2 has tested the efficiency of both approaches taken into account, the cross - correlation algorithm and the state of polarization algorithm. The results prove state of polarization to be more efficient. Let us consider the results of the state of polarization demultiplexer algorithm in a laboratory experiment setup.

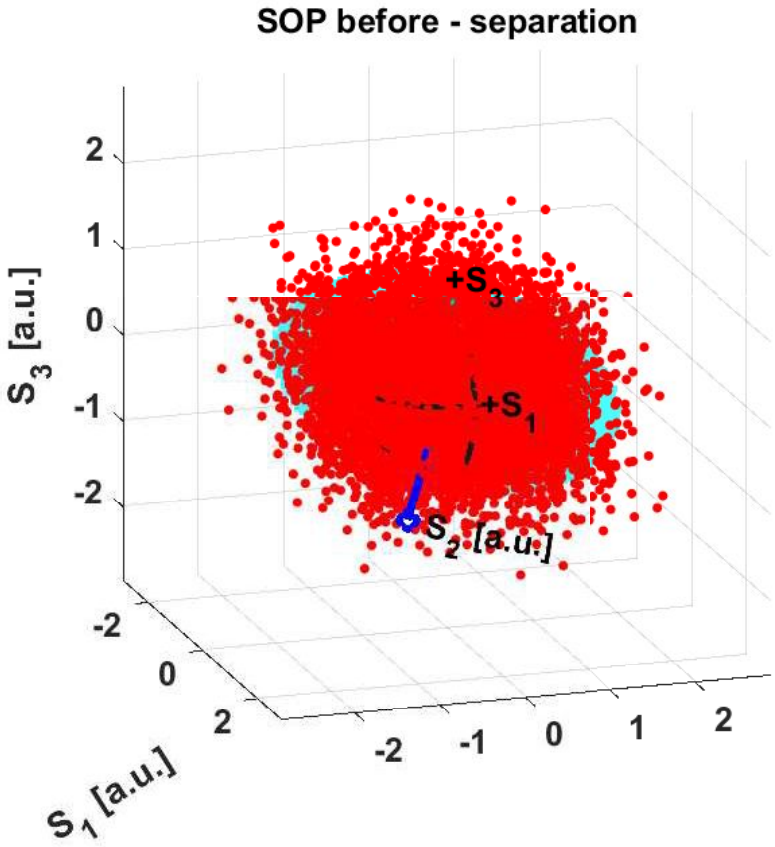


Fig 5.7: state of polarization before demultiplexer view 1.

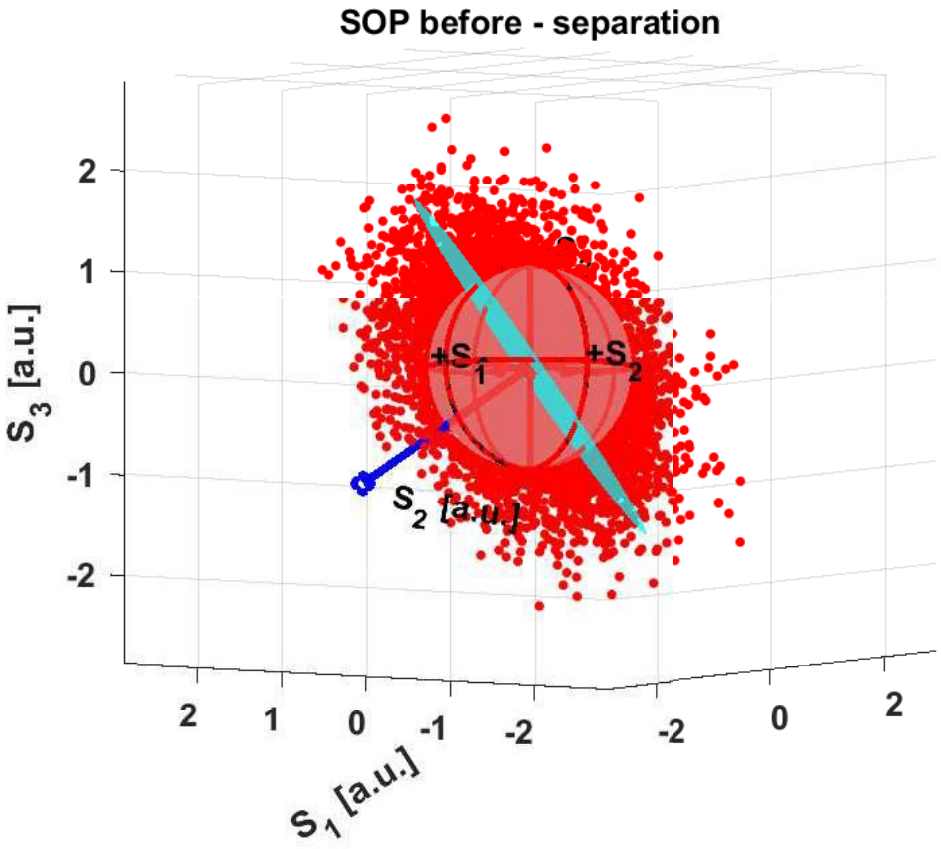


Fig 5.8: state of polarization before demultiplexer view 2.

The polarization demultiplexer is able extract the lens-like ellipsoid and its normal vector where all the state of polarization points concentrate. In the figure 5.8, it can be appreciated the lens-like ellipsoid and its normal vector painted in blue.

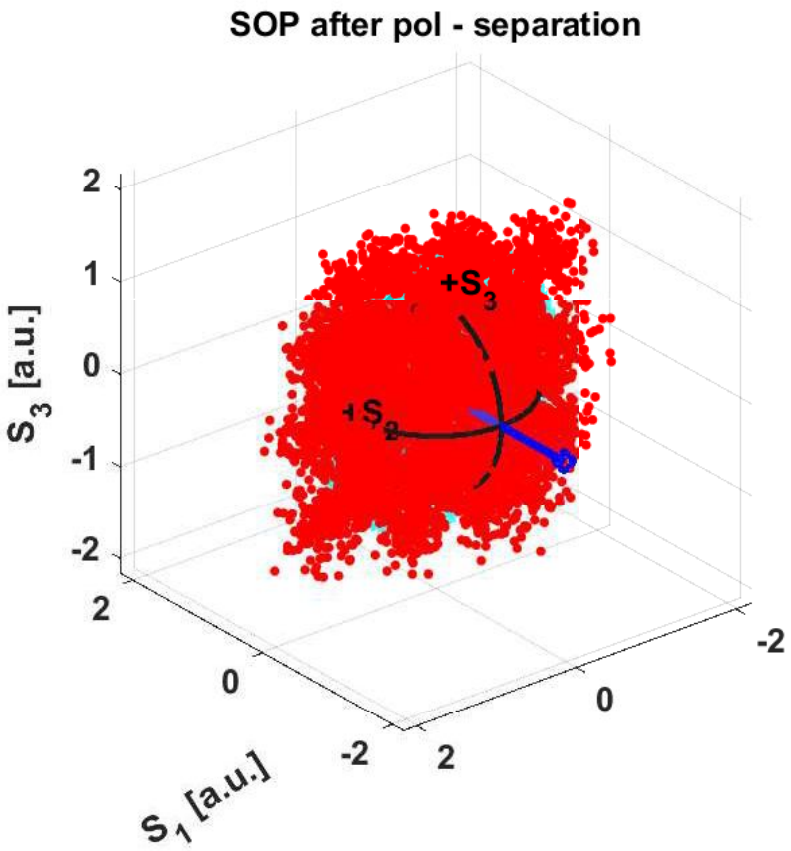


Fig 5.9: state of polarization after demultiplexer view 1.

The contribution of this thesis comes with the extraction of the phase angles and the appliance of the sequential rotation; first around the S_3 axis, then around the S_2 axis to finally lie the states of polarization ellipsoid into the S_2S_3 plane.

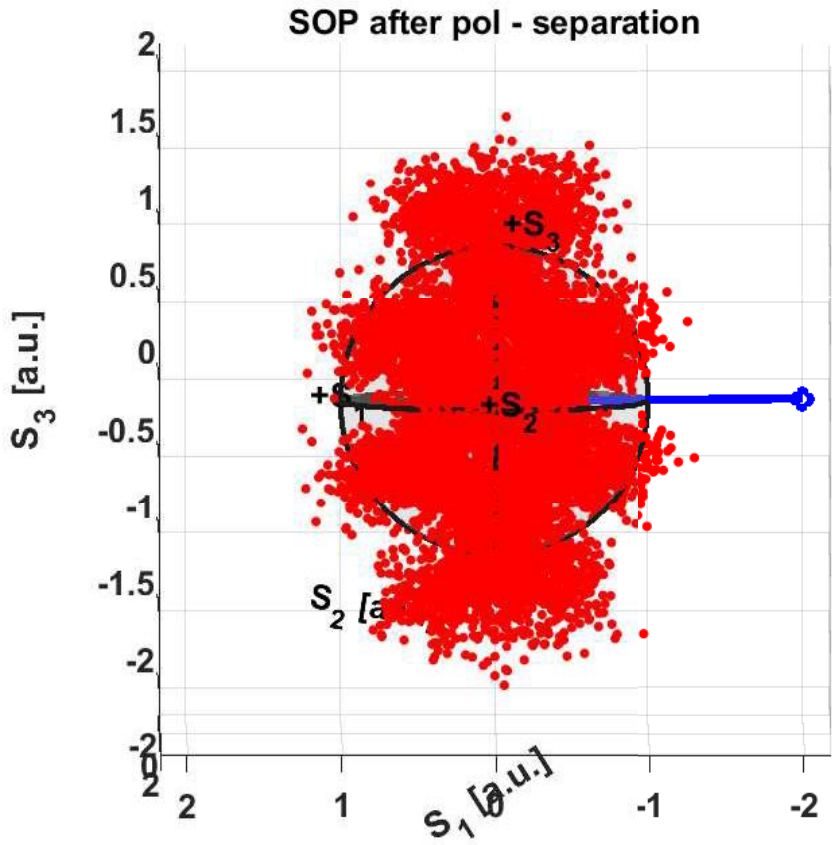


Fig 5.10: state of polarization after demultiplexer view 2.

The demultiplexer is able to rotate the len-like ellipsoid in a way in which the normal vector is pointing towards the S1 axis which represents the horizontal or vertical polarization. The figure 5.10 shows the decomposition of the mixed polarization **X**, **Y**. It is a elimination of the crosstalk between polarizations by rotation of a state of polarization SOP lens-like ellipsoid.

5.3.3 Clock Phase Recovery

Since there is not adaptive equalizer in the experimental setup, a clock phase recovery module is needed to extract the timing information from data to allow the receiving process to take place.

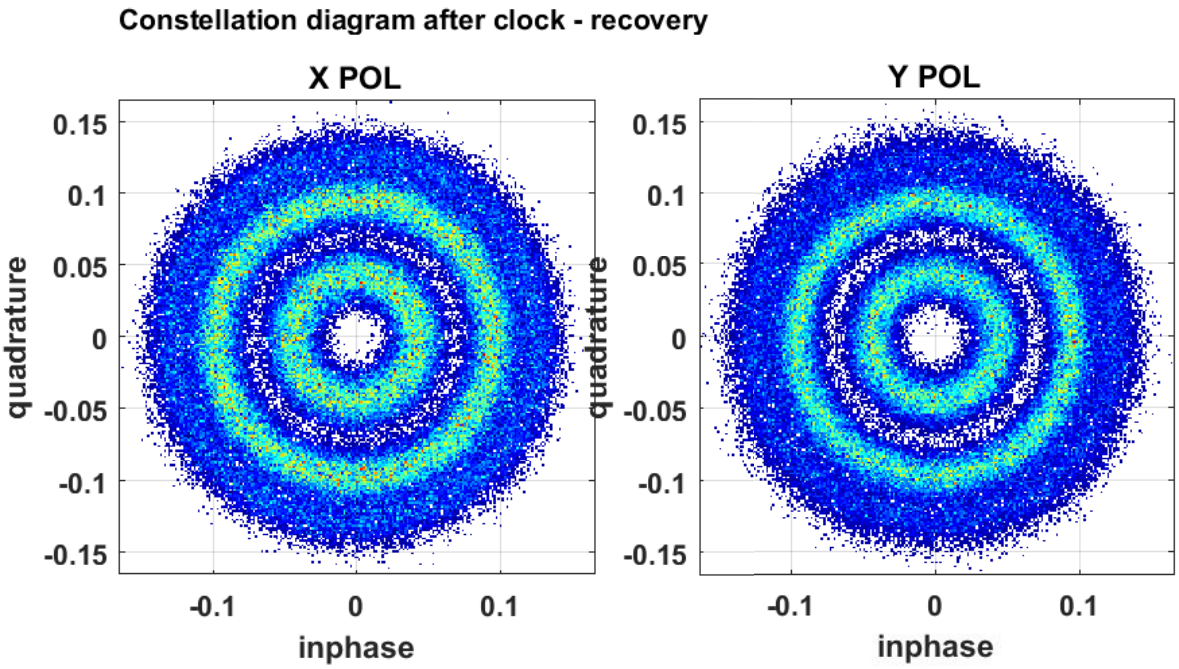


Fig 5.11: Constellation Diagram after clock phase recovery (1sa/s).

The figure 5.11 displays the three rings levels corresponding to the 16 - QAM energy levels. In this case, the constellation plot shows a signal at 1sa/s, the symbol samples are taken and the transition samples are left.

5.3.4 Carrier Frequency Offset

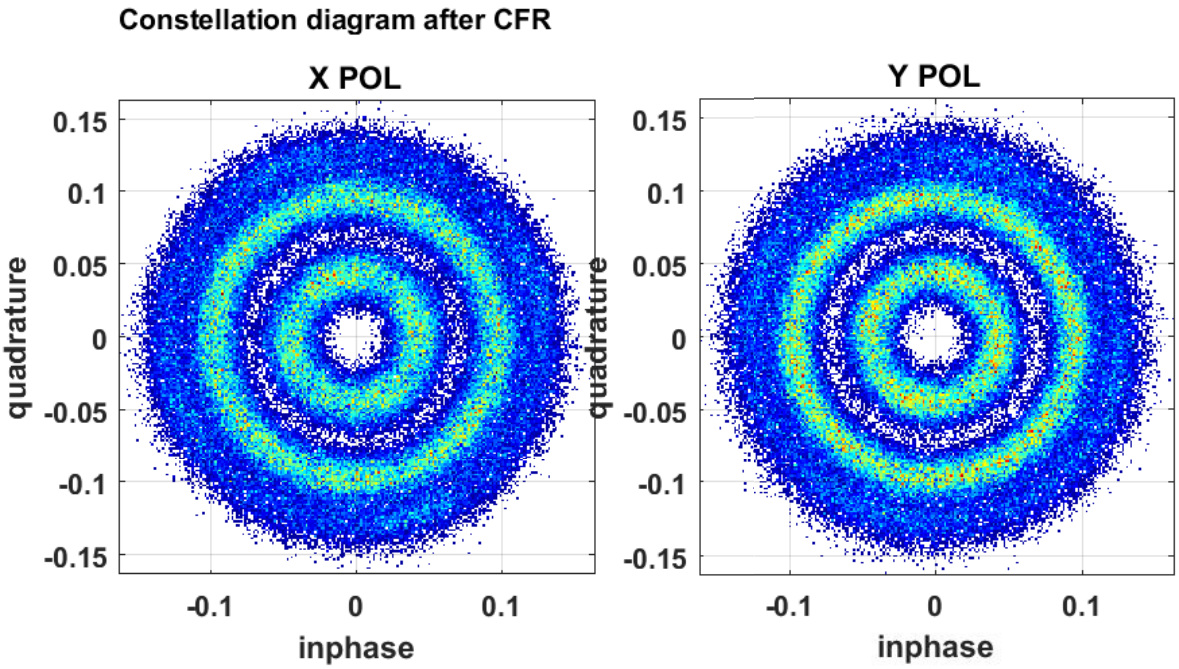


Fig 5.12: Constellation Diagram after carrier frequency offset (1sa/s).

Carrier frequency offset compensation removes rotations in phase of the constellation due to the frequency offset. Even though this effect cannot be seen in the constellation diagram.

5.3.5 Carrier Phase Recovery

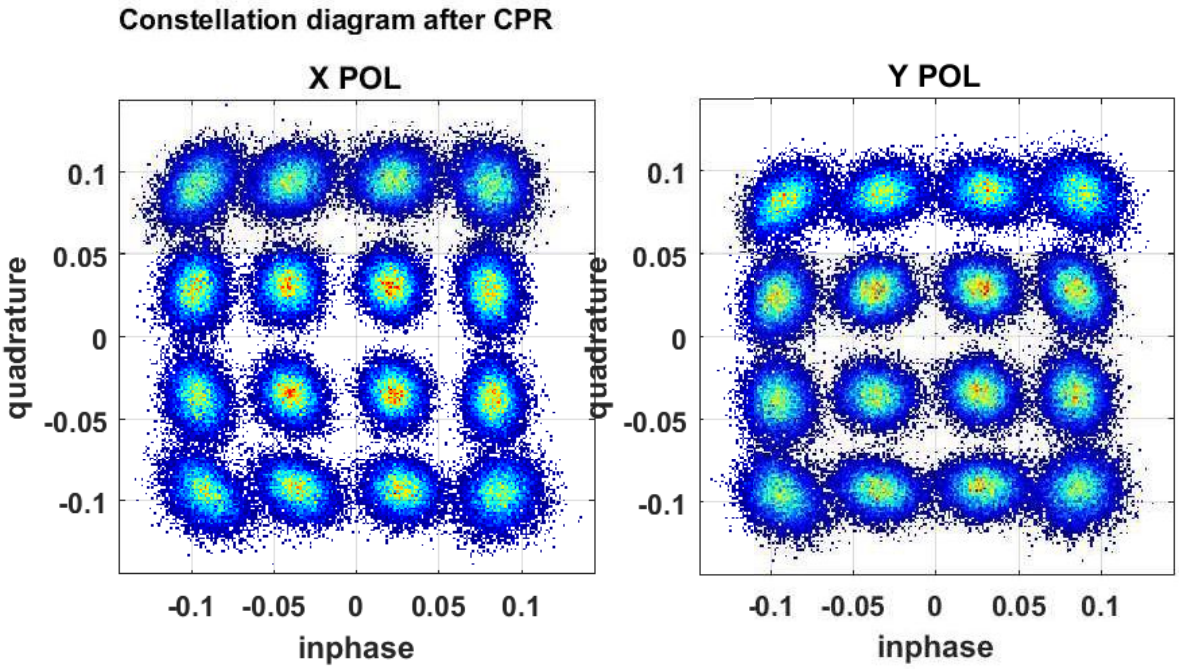


Fig 5.13: Constellation Diagram after carrier phase recovery (1sa/s).

After the carrier phase estimation most of the phase noise has been compensated. Finally, the cluttering that form the circular shape has been removed and only the actual points of the constellation are left.

The carrier phase recovery module has been modified to be able to work at two samples per symbol in order to run the setup at 32 Gsa/s as shown in the schematic (thesis contribution).

Among the several carrier phase recovery algorithms for optical coherent systems such as blind phase search, Viterbi sliding window and Viterbi block. The Viterbi sliding window approach has been applied in the digital signal processing of this thesis.

5.3.6 Bind Frame Synchronization

In order to estimate the kernel coefficients, both signals, the estimated signal and reference signal have to be synchronized, this is the goal of this module in the schematic.

The measured signal corresponds to the signal read in the oscilloscope and processed by the modified digital signal processing (bottom branch shown in the figure 5.2) and the reference signal corresponds to the signal loaded to the DAC.

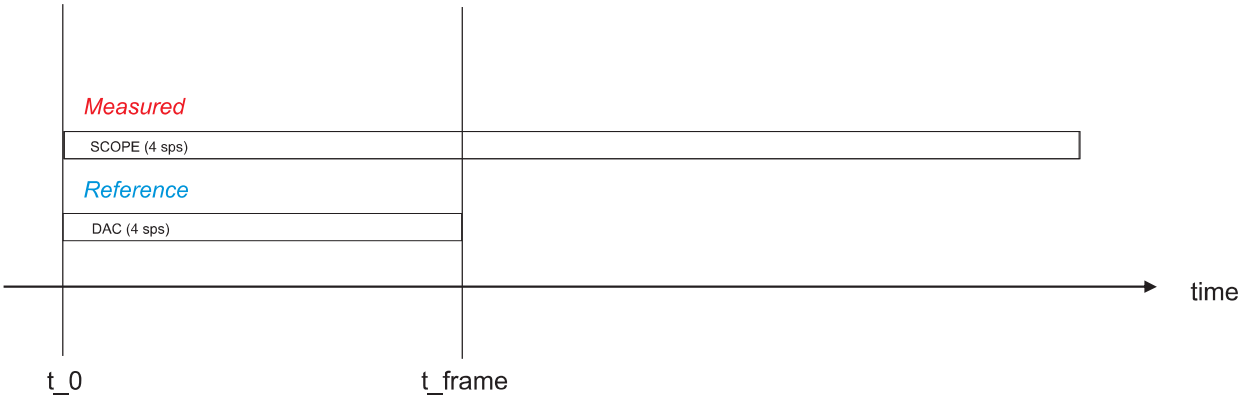


Fig 5.14: Blind frame synchronization scheme.

After the synchronization process, the starting time frame of both signals is common. Then, the signals are cropped to match the same frame size depending of the number of samples taken into account in the measurement process.

5.3.7 EVM Results

The error vector magnitude shows a good performance of the digital signal processing branch shown in the figure 5.2. The error vector magnitude (EVM) describe the vector errors between the received constellation and the ideal constellation. In this case, it is calculated as:

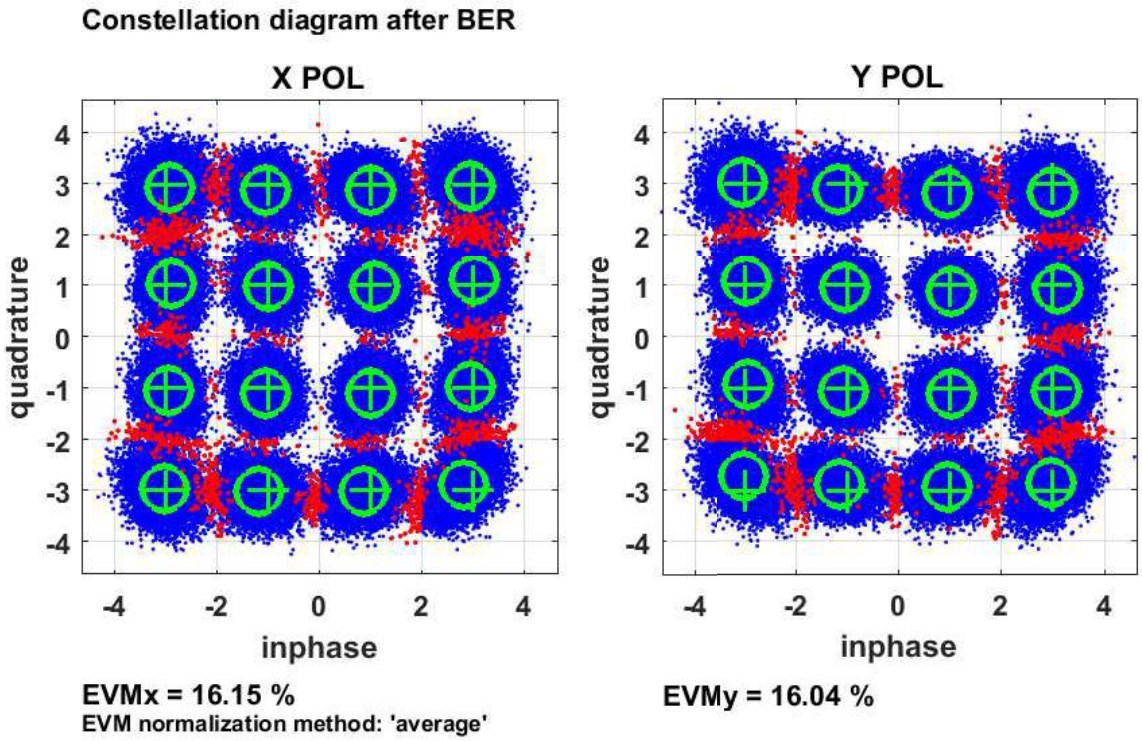


Fig 5.15: Constellation diagram.

The crosses plotted in the figure 5.16 represent the ideal position of the constellation points, while the circles represent the centers of gravity. It can be seen that all the crosses lay inside the circles, resulting in a EVM average parameter of **16%** which means a decent recovery through dsp.

5.4 Results

This is experiment procedure:
We want to receive a signal transmitted without digital pre-distortion (measured), the receiver should not have adaptive equalizer. Then the Kernel coefficients of the channel are estimated by using the transmitted signal, the received signal (measured) and the Volterra memory.

After identifying the Kernel coefficients of the system, the same transmitted signal is pre-distorted using the already identified Kernel coefficients, thus a new received signal is generated (simulated). Finally, the performance is tested by comparison of the simulated signal and the previously measured signal for each component I_x, Q_x, I_y and Q_y, this are the results:

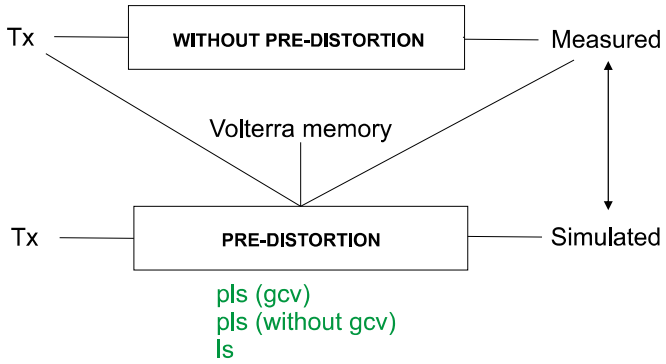
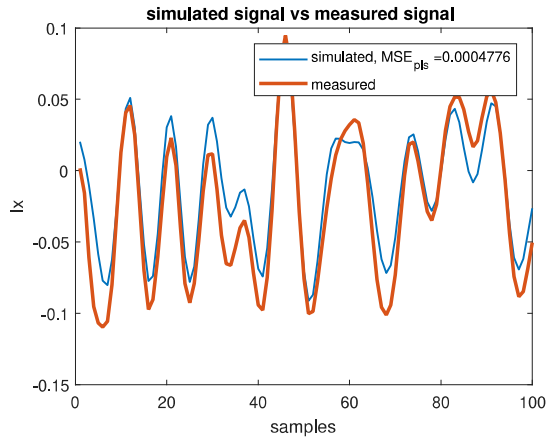
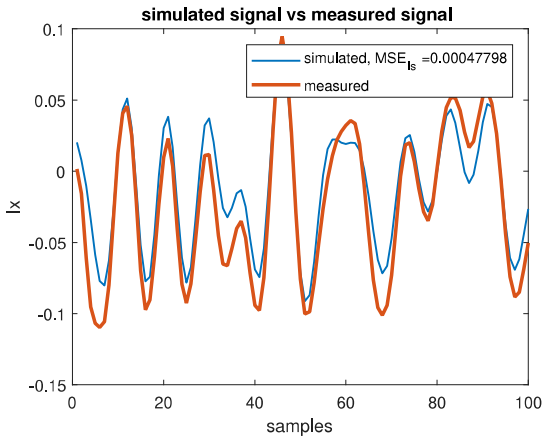
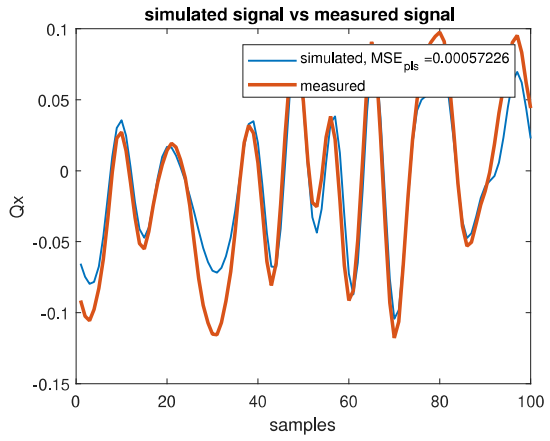


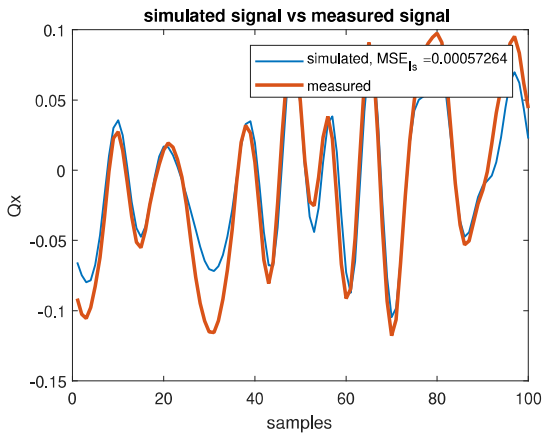
Fig 5.16: Experiment procedure schematic.

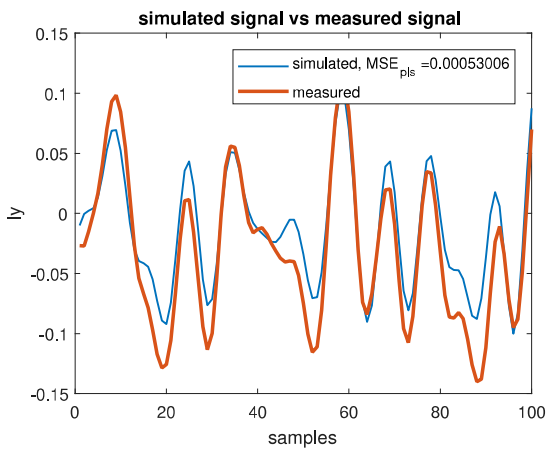
The setup of the experiment consists in the following system parameters:
- modulation format: **PDM - 16QAM**
- pulse shape: **Root Raised Cosine**
- roll-off: **0.1**
- laser wavelenght: **1550.1 nm / 193.4 THz**
- number of symbols: 2^{15} **symbols (DAC)**
- number of samples: 2^{17} **samples (DAC)**
- reduced number of samples: $\approx 2^{15}$ **samples**
- symbol rate: **16 GBaud**
- oscilloscope IDN: **TEKTRONIX, DPO72004, B030945**
- channel bandwidth: **20.0 GHz**
- channel deskew: **0.00, -0.02, 0.05, -0.021 ns**

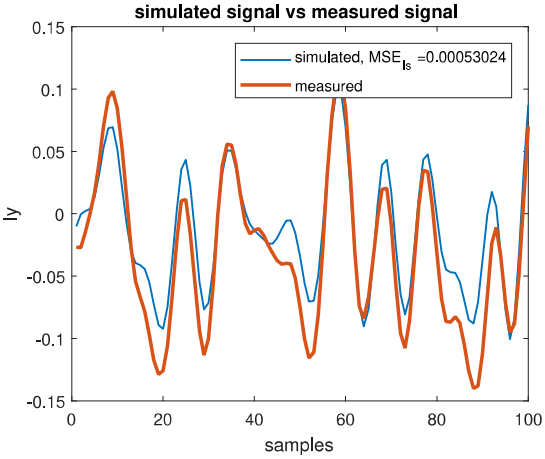

simulated signal vs measured signal
simulated, MSE$_{pls}$ =0.0004776
measured
Ix
samples

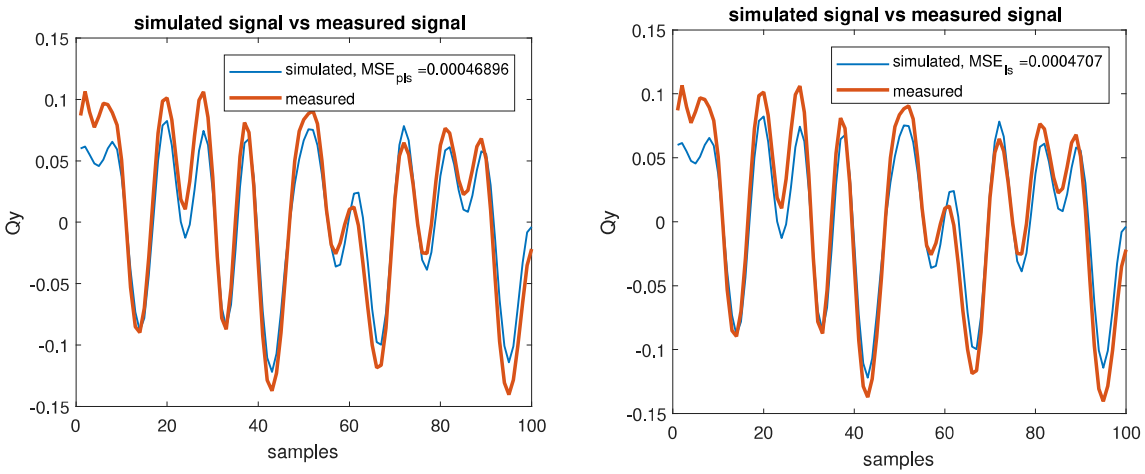

simulated signal vs measured signal
simulated, MSE$_{ls}$ =0.00047798
measured
Ix
samples


simulated signal vs measured signal
simulated, MSE$_{pls}$ =0.00057226
measured
Qx
samples


simulated signal vs measured signal
simulated, MSE$_{ls}$ =0.00057264
measured
Qx
samples


simulated signal vs measured signal
simulated, MSE$_{pls}$ =0.00053006
measured
Iy
samples


simulated signal vs measured signal
simulated, MSE$_{ls}$ =0.00053024
measured
Iy
samples

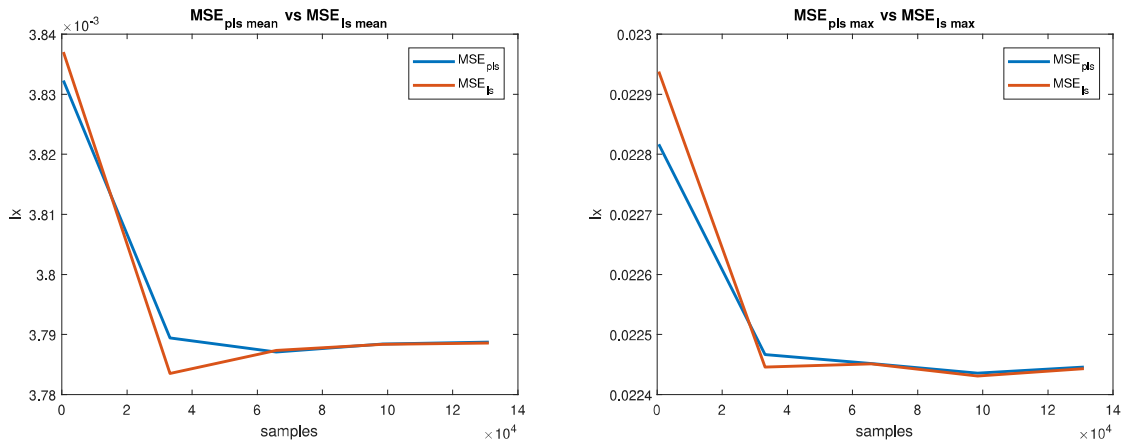


(a) (b)

Fig 5.17: Experiment results for penalized least squares (pls) (a) and least squares (ls) (b).

The previous figure 5.17 shows the error parameter MSE between the simulated signal and the measured signal while using the penalized least squares approach figure 5.17 (a) with generalized cross validation and filter operation mode (section 4.2) and the least squares approach figure 5.17 (b) for the Kernel coefficients identification.

As it is explained in section 3.3.1, the ordinary least squares is an linear inverse problem, under a small amount of samples it does not converge to a unique solution. In order to test the performance of the penalized least squares in comparison with the least squares, different experiments have been taken into account while varying the number of samples for estimation.



(a) (b)

Fig 5.18: Experimental results for different number of samples estimation. Error parameters: mean squared error MSE (a), maximum squared error (b).

The comparison consists in estimating the mean squared error figure 5.18 a) and maximum squared error figure 5.18 b) MSE for the penalized least squares against the ordinary least squares. The experimental results show a slight advantage in the performance of the penalized least squares, especially in the case of very low number of samples used in the estimation.

Even though the positive results of the simulations, for this particular case, the penalized least squares approach has not shown a significant advantage in performance in comparison with the least squares approach.

Chapter 6

Conclusion

6.1 Context

In the market of coherent optical communications systems, it is wildly common the adoption of polarization division multiplexing (PDM) assisted by digital signal processing (DSP). One of the current problems and main limitation in the coherent optical fiber transmission are the nonlinear effects.

The proposed solution aims to diminish the nonlinear effects of the transmitter component in the communication system by pre-distorting the signal at the coherent transmitter. Opening the possibility in this way of achieving reliable higher capacity communications.

6.2 Problem

The increasing demand of higher capacity leads to higher spectral efficiency. One way to achieve this is through high cardinality modulations. However, the use of such modulation formats implies changes in the optical system.

Changes such as employing coherent detection and digital signal processing (DSP) in the optical receiver in order to meet the quality component requirements for high cardinality modulations.

The digital signal processing chain needs a further improvement, in particular the digital nonlinear pre-distortion. The performance of the different algorithms applied for digital

nonlinear pre-distortion (ordinal least squares) do not perform well under short data set or/and under low SNR regime.

In order to soften the quality components requirement of high cardinality modulations, the performance of current digital nonlinear pre-distortion must be optimized.

6.3 Thesis Approach

Literature provides us with different solutions to this problem. One such solution is the use of penalized least squares, which is based on adding a term of penalty or smoothness to the previous approach (ordinary least squares) [13]. This regularization or penalization function contains *a priori* information of the signal.

The thesis contribution consists in the design and Matlab implementation of a regularization algorithm for the penalized least squares approach. The thesis proposes a regularization method based on the spectrum shape of the input signal or the filter shape used in the input signal generation to be able to acquire good estimations under noisy and short data in coherent receiver.

Thanks to the spectrum shape of the signal, the regularization method can penalize those components that do not correspond to the signal.

The approach has been tested for a **PDM- 16QAM**, pulse shape **root raised cosine**, roll - off **0.1**, bandwidth **20 GHz** at **16 GBauds** while using the **CFP2 - ACO** transceiver property of the HHI.

In addition, a generalized cross validation has been implemented in Matlab to provide an optimum regularization parameter [13] [14].

6.4 Validation procedure

The validation procedure of the implemented algorithms by this thesis is the following: First, a simulation of the Matlab algorithms in a control environment with arbitrary nonlinearities is implemented. Then, an experimental setup at the HHI laboratory is used in order to obtain empirical results.

Simulation

As an initial test for the validity of the algorithms proposed in this thesis, the following Matlab simulations were considered. First, a simulation testing the performance of the regularization method implemented in the penalized least squares for different bandwidth and roll-off values of the input signal spectrum shape.

Next, a simulation verifying the improvement of the generalized cross validation in the penalized least squares. The simulation tests the mean squared error MSE obtained by both approaches in a large SNR range. Then, a simulation checking the improvement of the penalized least squares against the ordinary least squares. The analyzed error parameter is the mean squared error MSE in regard to SNR.

Finally, a simulation comparing the performance of the two studied approaches for the polarization demultiplexer (cross - correlation and state of polarization (SOP)). The error parameter evaluated is the vector error magnitude EVM with respect to SNR.

Experiment

The experiment at the HHI laboratory give us insights about the observed performance. First, the experiment aims to test the performance of the penalized least squares against the ordinary least squares while providing factual results.

For each signal component (I_X, I_Y, Q_X, Q_Y), a performance comparison between the observed signal without digital pre-distortion and the pre-distorted signal by both methods is presented.

Lastly, the same experiment setup is taken into account while varying the number of samples of the data set for the comparison between the two methods estimation.

6.4.1 Results

The results from the experiment setup show a performance comparison found on the mean squared error MSE, the comparison between the pre-distorted signal by the kernel coefficients and the measured non-predistorted signal in the lab. It does not show improvement in the performance of the penalized least squares in contrast to the least squares.

The results have been used to compare the performance of the penalized least squares in opposition to least squares algorithm with a varying number of samples per estimation. Once more, the results show a minor enhancement in the penalized least square approach.

To conclude, even though the simulation results were promising, the experimental result does not present a convincing improvement in this specific case.

6.5 Results Discussion

As a conclusion, the laboratory results do not reveal a significant improvement in the case of penalized least squares against ordinary least squares. It proves that the performance of the ordinary least squares is good enough even in short data set or even in low signal to noise ratio regimes (SNR).

6.6 Summary

This thesis goal was the design and Matlab implementation of a regularization method in the penalized least squares approach. The design of the regularization algorithm consists in the information of the input signal shape in order to penalized the components foreing to the signal.

As a form of testing the regularization algorithm, a set of simulations and experiment have been deployed. As an overall conclusion, the simulation results of the regularization algorithm in penalized least squares for a simulated controlled environment with arbitrary nonlinearities were very favorable.

The simulation results show how the implemented regularization technique in the penalized least squares clearly outperform the ordinary least squares. However, in the case of the experiment in the HHI laboratory, the empirical results of the regularization method do not seem to have such a positive impact for this specific scenario.

Further study must be considered.

6.7 Future research

The research of this thesis can be extended, the implementation of adaptive regularization algorithms for least squares could be considered. Even though, the literature on adaptive regularization is rather limited, due to the fact that generic time-variant regularization leads to cubic computational complexity.

Chapter 7

Appendices

7.1 Appendix A - Mach - Zehnder Modulator (MZM)

Phase modulation depends on the wavelength, the electrode length l_{el} and the change in the effective refractive index Δn_{eff}. The following figure shows a phase modulator, where the back blocks represent the electrode (voltage u(t)) and the orange block represents the waveguide.

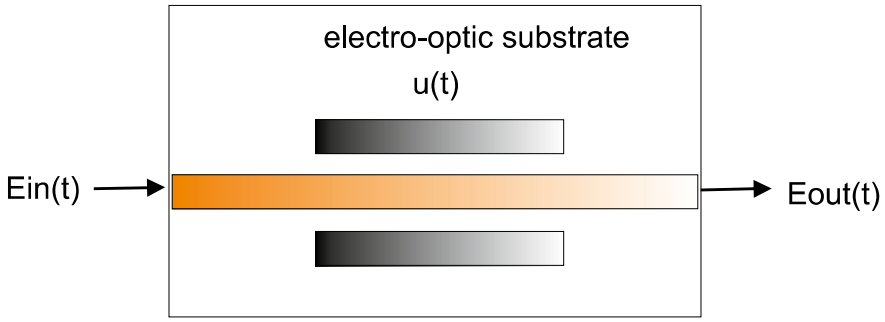


Fig 7.1: Phase Modulator

There is a need for the specification of the parameter V_π which corresponds to the necessary voltage to produce a phase shift of π.

$$\varphi(t) = \frac{\lambda}{2\pi}\Delta n_{eff}.l_{el} \sim u(t)$$

$$E_{out}(t) = E_{in}(t)e^{j\varphi(t)} = E_{in}(t)e^{j\frac{u(t)}{V_\pi}\pi}$$

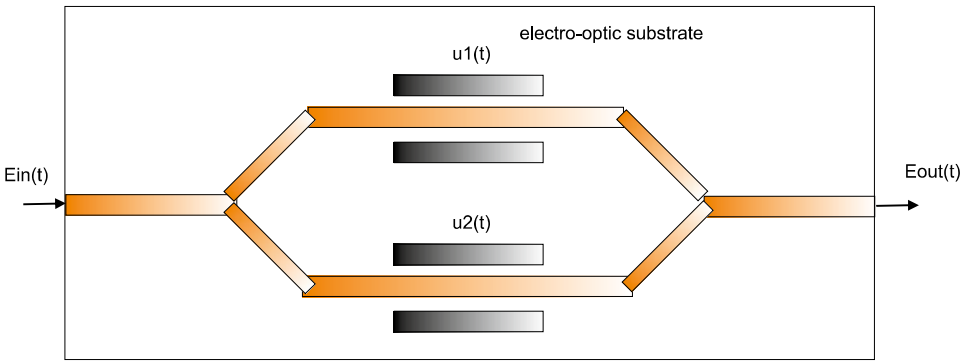


Fig 7.2: Mach - Zehnder Modulator (MZM)

Then, two phase modulators can be marged in parallel while using an interferometric structure. The input light is divided into two branches, the branches experience different phase shifts. Finally, the two paths are recombined.

The result is a range of constructive and destructive interferences.

$$\frac{E_{out}(t)}{E_{in}(t)} = \frac{1}{2}(e^{j\varphi_1(t)} + e^{j\varphi_2(t)})$$

7.2 Appendix B - Computation of the Matrix T_p

In the regularization process of the penalized least squares. In order to compute the matrix T, the position of the different permutations of an specific index have to be found. The approach of this algorithm is to use the sum of the elements of the index in order to identify the permutation of that index, for instance:

$$\left(h_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 2 & 2 \end{bmatrix} \right) \left(\hat{h}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 0 & 0 \\ 1 & 2 \\ 2 & 0 \\ 2 & 1 \\ 0 & 0 \end{bmatrix} \right) \left(SUM = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \\ 0 \\ 3 \\ 2 \\ 3 \\ 0 \end{bmatrix} \right)$$

In order to avoid confusing permutations such as [1,1] = 2 and [2,0] = 2, in case of having the two equal elements of the index, the sum of them will be zero as presented in the second matrix above. Then the sum of the elements will be computed as in the third matrix.

7.3 Appendix C - Linear inverse problems

Literature defines an linear inverse problem as a problem which does not meet all the requirements to be considered a well - posed problem.

$$Au = f \tag{7.1}$$

$$A \in \mathcal{L}(H, F) \tag{7.2}$$

Where H,F correspond to Hilbert spaces.

The requirements for a well - posed problem are:

1) for all f $\in \mathcal{F}$ has solution u_* $\in$ H.
2) for all f $\in \mathcal{F}$ solution is unique.
3) solution u_* depends continuously on the data.

If one of the three previous requirements is not satisfied, then the presented problem is an linear inverse problem.

Literature

[1] Kikuchi, K. (2010). Coherent optical communications: Historical perspectives and future directions. High Spectral Density Optical Communication Technologies, 11-49.

[2] Ip, E., Lau, A. P. T., Barros, D. J., Kahn, J. M. (2008). Coherent detection in optical fiber systems. Optics express, 16(2), 753-791.

[3] Dwivedi, N., Bohara, V. A. (2014). Fixed-point digital predistortion system for nonlinear high power amplifiers (Doctoral dissertation, IIIT Delhi).

[4] Stolen, R. H.; Ashkin, A. Optical Kerr effect in glass waveguide. Applied Physics Letters, 1973, vol. 22, no 6, p. 294-296.

[5] Fludger, C. R., Kupfer, T. (2016). Transmitter impairment mitigation and monitoring for high baud-rate, high order modulation systems. In ECOC 2016; 42nd European Conference on Optical Communication; Proceedings of (pp. 1-3). VDE.

[6] Savory, S. J. (2010). Digital coherent optical receivers: Algorithms and subsystems. IEEE Journal of Selected Topics in Quantum Electronics, 16(5), 1164-1179.

[7] Ip, E., Lau, A. P. T., Barros, D. J., Kahn, J. M. (2008). Coherent detection in optical fiber systems. Optics express, 16(2), 753-791.

[8] Yoshino, M., Haruyama, S., Nakagawa, M. (2008). High-accuracy positioning system using visible LED lights and image sensor. In Radio and Wireless Symposium, 2008 IEEE (pp. 439-442). IEEE.

[9] Komine, T., Nakagawa, M. (2004). Fundamental analysis for visible-light communication system using LED lights. IEEE transactions on Consumer Electronics, 50(1), 100-107.

[10] Mesleh, R., Elgala, H., Haas, H. (2012). LED nonlinearity mitigation techniques in optical wireless OFDM communication systems. Journal of Optical Communications and Networking, 4(11), 865-875.

[11] Schetzen, M. (1980). The Volterra and Wiener theories of nonlinear systems.

[12] Berenguer, P. W., Nölle, M., Molle, L., Raman, T., Napoli, A., Schubert, C., Fischer, J. K. (2016). Nonlinear digital pre-distortion of transmitter components. Journal of Lightwave Technology, 34(8), 1739-1745.

[13] Nowak, R. D. (1998). Penalized least squares estimation of Volterra filters and higher order statistics. IEEE Transactions on signal processing, 46(2), 419-428.

[14] Ueki, M., Fueda, K. (2010). Optimal tuning parameter estimation in maximum penalized likelihood method. Annals of the Institute of Statistical Mathematics, 62(3), 413-438.

[15] Schaffrin, B. (2008). On penalized least-squares: Its mean squared error and a quasi-optimal weight ratio. Recent Advances in Linear Models and Related Areas, 313-322.

[16] Thomsen, B. C., Maher, R., Millar, D., Savory, S. (2011,). Burst mode receiver for 112 Gb/s DP-QPSK. In European Conference and Exposition on Optical Communications (pp. Mo-2). Optical Society of America.

[17] Millar, D. S., Savory, S. J. (2011). Blind adaptive equalization of polarization-switched QPSK modulation. Optics express, 19(9), 8533-8538.

[18] Savory, S. J. (2010). Digital coherent optical receivers: Algorithms and subsystems. IEEE Journal of Selected Topics in Quantum Electronics, 16(5), 1164-1179.

[19] Szafraniec, B., Nebendahl, B., Marshall, T. (2010). Polarization demultiplexing in Stokes space. Optics Express, 18(17), 17928-17939.

[20] Malykin, G. B. (1997). Use of the poincare sphere in polarization optics and classical and quantum mechanics. review. Radiophysics and quantum electronics, 40(3), 175-195.

[21] Czegledi, C. B., Karlsson, M., Agrell, E., Johannisson, P. (2016). Polarization drift channel model for coherent fibre-optic systems. Scientific reports, 6, 21217.

[22] Khanna, G., Calabro, S., Spinnler, B., De Man, E., Hanik, N. (2015). Joint adaptive pre-compensation of transmitter I/Q skew and frequency response for high order modulation formats and high baud rates. In Optical Fiber Communications Conference and Exhibition (OFC), 2015 (pp. 1-3). IEEE.

[23] Khanna, G., Spinnler, B., Calabro, S., De Man, E., Hanik, N. (2016). A robust adaptive pre-distortion method for optical communication transmitters. IEEE Photonics Technology Letters, 28(7), 752-755.

[24] Khanna, G., Spinnler, B., Calabro, S., De Man, E., Chen, Y., Hanik, N. (2017). Adaptive Transmitter Pre-Distortion using Feedback from the Far-End Receiver. IEEE Photonics Technology Letters.

[25] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.

[26] Tibshirani, R., Wainwright, M., Hastie, T. (2015). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.

[27] Zhang, T. (2004, July). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In Proceedings of the twenty-first international conference on Machine learning (p. 116). ACM.

[28] Randolph, T. W., Zhao, S., Copeland, W., Hullar, M., Shojaie, A. (2015). Kernel-Penalized regression for analysis of microbiome data. arXiv preprint arXiv:1511.00297.

[29] Krivobokova, T. (2006). Theoretical and practical aspects of penalized spline smoothing.

[30] Brewer, J. (1978). Kronecker products and matrix calculus in system theory. IEEE Transactions on circuits and systems, 25(9), 772-781.