Complex-Valued Signals: a Tutorial for Non-Circular Data Processing and Analysis

Rafael de Carvalho Bluhm, João César Moura Mota and Charles Casimiro Cavalcante

***Abstract*—This work explores the versatility of complex-valued signals present in data representation over diverse applications. Integrating complex values into signals and processing structures proves powerful in system modeling, demonstrating effectiveness in many scenarios. The inherent mathematical efficiency of complex numbers enhances both elegance and practical utility, simplifying calculations and expanding capabilities. However, the use of complex notation introduces considerations related to the intrinsic geometry of the data. This tutorial unravels the theoretical foundations and practical applications in adaptive filtering of complex-valued signals, emphasizing the importance of understanding the displayed profiles of the studied signals and data. Using complex augmented notation and widely linear processing, superior performance is attainable in specific cases, providing valuable information to readers interested in signal processing.**

***Index Terms*—Complex-valued signals, widely linear, improper, augmented.**

I. Introduction

COMPLEX-valued signals, characterized by their real and imaginary components (or magnitude and phase), provide a robust and versatile approach to data representation in diverse application domains. These domains cover a broad spectrum, including biomedical data processing, radar systems, mobile communications, wind modeling, quantum mechanics, and beyond [1]–[6]. The integration of complex values into signals and their processing structures has become a powerful tool in signal and system modeling, showcasing effectiveness in various applications in the field of data and signal processing [7]–[11].

This approach extends beyond the realm of data analysis and has been influential since the development of complex numbers. Complex numbers algebra and calculus play a fundamental role in simplifying intricate integrals, modeling fluid flow in dynamics, and designing transmission lines, among many other applications [12]–[15]. The curious reader might wonder why complex values are used instead of treating their real and imaginary components separately. The answer lies in their rich algebraic structure, which streamlines calculations while maintaining concise and elegant notation. This structural advantage not only saves researchers time and effort, but also broadens the analytical possibilities beyond those offered by real numbers alone. As such, the mathematical efficiency of complex numbers becomes a key factor in both the theoretical elegance and practical effectiveness of the methods in which they are employed [16].

The use of complex notation brings several advantages, but it also introduces a new factor to be considered: the intrinsic geometry of the data, which comes from the representation in two components [17], [18]. As will be seen in this work, the profile of the data (the distribution of the data over the complex plane) concerning its real and imaginary components can significantly influence the performance in signal processing.

A. Road-map to the present

Knowledge is inherently dispersed, as it gradually accumulates through the efforts of various researchers across different locations and time periods. This dispersion makes it challenging to trace the origin of specific ideas. In this work, we begin our exploration by tracking the definition of the complex Gaussian distribution, a concept that is crucial for the analysis of circularity and complex adaptive filtering. One of the first discussions on this topic can be found in [19]. These ideas were later applied to statistics for the first time in [20], and since then, numerous studies across various fields have further developed this topic.

The main point, and the focus of the present work, is the treatment of non-circular distributions in signal processing, which has not been addressed in previous works despite being one of their principal tools. Only in [21] is perhaps the first mention of the term *proper*. This work deals with complex processes and shows that the maximum differential entropy is attained for a complex vector if it is “proper” and zero-mean Gaussian. The article also defines the term *circular stationarity* of a process.

One of the seminal works on the study of circularity properties (the ratio aspect between the real and imaginary components of the complex signal) is done in [22]. This study concludes that classical estimation methods must naturally handle circular vectors. The paper also suggests that, to achieve better results, the theory should be modified to account for cases where the signal vectors are not circular. The subsequent work, by the same author, in [23], utilizes the definition of widely linear transformations to estimate complex signals, considering second-order statistics. This approach, used to this day, is the simplest method that successfully captures all the necessary second-order statistics (correlation and the so-called pseudo-correlation, as will be discussed

Rafael de Carvalho Bluhm (ORCID: 0009-0003-4942-1535), João César Moura Mota (ORCID: 0000-0001-9890-2982) and Charles Casimiro Cavalcante (ORCID: 0000-0002-4198-4064) are with Federal University of Ceará (UFC), Fortaleza-CE, Brazil, E-mails: faelbluhm@alu.ufc.br, mota@gtel.ufc.br, charles@gtel.ufc.br.

The authors thank to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001 and CNPq (Procs. 308512/2023-5 and 409228/2021-4) for partial financial support.

Submission: 2025-05-29, First decision: 2025-08-04, Acceptance: 2025-09-09, Publication: 2025-10-04.

Digital Object Identifier: 10.14209/jcis.2025.7

later). The same work represents an attempt to adapt the theory that was previously missing in the resolution related to estimation. The author also published additional important articles on the topic in [24], [25].

In [26], a comprehensive description of circular (and non-circular) distributions is provided geometrically. This work shows that the circularity coefficient, which is based on second-order statistics, has a direct geometric interpretation: the magnitude is related to the eccentricity and the phase to the angle of the complex Gaussian ellipse. Since then, highly regarded books have been published on the subject of complex signals and the consideration of non-circularity in the design of adaptive structures or neural networks, as seen in [27]–[29].

Many other publications have been produced since then, as can be seen in [30]–[34]. Fig. 1 shows the timeline of some of these publications in chronological order.

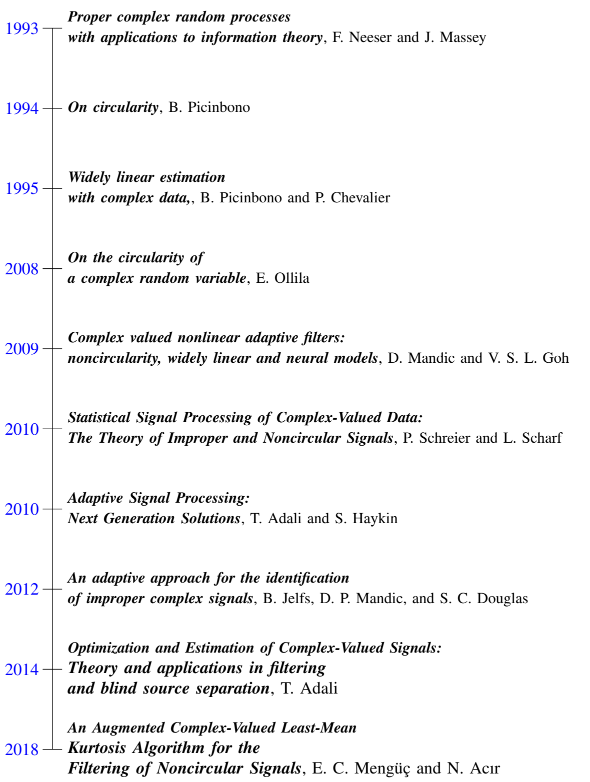


Fig. 1. Timeline for some publications related to the topic.

B. Objectives of this study in complex signal processing

Building on these considerations, this tutorial explores the essential theoretical foundations and practical applications of adaptive filtering in the field of complex numbers, considering the complex structuring and operation of data that can be represented in this form. Our goal is to clarify the challenges associated with this type of processing, highlighting its significance in extracting meaningful information from the data. We will demonstrate that by using a specialized notation (complex augmented) to represent the data, along with a specific processing structure (widely linear), superior performance can be achieved compared to conventional complex filtering.

This tutorial is structured to guide the reader through the basic fundamentals of complex signal processing and some adaptive techniques for the effective analysis of complex signals. The techniques described here can be applied to other constructions or models of particular interest to the reader. Fig. 1 shows a timeline that presents various books and articles related to the theme published in the correlated areas of signal processing. This timeline was constructed based on key articles recommended for reading leading up to the current work and does not imply any specific preference or ranking on the part of the author. For a more comprehensive and extensive list, the reader can and should refer to the bibliography at the end of this document.

C. Organization

This work is structured to provide a comprehensive introduction to discrete signal processing, with a focus on non-recursive adaptive filters and a special emphasis on processing complex signals with distinct profile characteristics. To establish the foundational concepts of filtering and signals, Sections Sections II and III explore the fundamentals of discrete linear filters and their typical applications. This introduction serves as the basis for the comparisons made throughout the paper, where the design of conventional filters will be contrasted with an alternative approach that considers the specific characteristics of the signal profile.

In Section IV, we introduce the concept of Wirtinger calculus, a fundamental algebraic and calculus tool used when dealing with complex variables. This method employs a value and its conjugate as variables, along with derivatives with respect to these variables. It is crucial for the reader to comprehend this section, as the notation throughout the paper is based on it, facilitating the algebraic manipulation of complex expressions. At the end of the section, the optimal solution for the design of filters developed in the preceding section using this notation is presented.

Section V explores the concept of circularity (or its absence), which is a broader notion, as well as that of propriety/impropriety, a more specific definition based exclusively on second-order statistics. These terms are intrinsically related to the complex profile of the signal, categorizing them into two essential groups throughout the development of the work: signals referred to as proper and signals referred to as improper. The second-order statistics characterizing complex signals include covariance, a concept already employed in the usual complex optimal solution (Section IV-F), and pseudo-covariance, which must be taken into account when the signal is said to be improper. In the conclusion of the section, it is detailed how the signals used in simulations will be modeled considering the signal profile.

Section VI presents the design of a discrete transversal filter capable of taking into account both covariance and pseudo-covariance. This design is based on an expression called widely linear, hence the origin of its name. This filtering process is based on the linear combination of the signal with its conjugate value.

Section VII emphasize some of the fundamental numerical tools applied in linear and widely linear processing, aiming to achieve optimal solutions iteratively. Recursive algorithms are expressed in complex notation and subsequently applied in simulations in Section IX. The objective is to make a comparison between linear and widely linear methods in the realm of discrete processing, considering the mean square error in the design of discrete adaptive filters. This work concludes with the insights drawn from the simulations presented in this section.

As this work concludes, it seeks not only to provide valuable findings but also to engage and captivate the reader. By introducing these insights, we invite the reader to explore the research, either to expand their knowledge in the field or simply out of curiosity. We hope that this article sparks interest and encourages readers to join and contribute to the ongoing exploration and advancement of this area of study.

D. Notation

Before examining the main topics of the tutorial, it is advisable to cover a brief section on the notation used. We have opted for a notation commonly employed in books and scientific articles in the signal processing field. In summary, the following scheme has been employed:

- Deterministic scalars are written in lowercase, without bold.
- Random scalars and Z-transforms are written in uppercase without bold.
- Vectors are represented in lowercase bold.
- The matrices and vectors of the random variables are represented in bold uppercase letters.
- The letter "j" is reserved for the imaginary unit, that is, $j = \sqrt{-1}$. The operation $(\cdot)^*$ denotes the conjugation, where, for $z = x + jy$, its conjugate is $z^* = x - jy$. This conjugation operation can be performed element-wise on vectors and matrices.
- $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^{-1}$ represent the operations of transposition, transposition plus conjugation, and inversion, respectively.
- The modulus $|\cdot|$ of a complex number $z = x + jy$ is given by $|z|^2 = zz^*$.
- Augmented vectors and matrices, as will be seen, are denoted by an underline below the corresponding letter.
- $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{F}$ represent the sets of integers, real numbers, complex numbers, and some arbitrary field, respectively.

Examples:

- Scalar: x.
- Z-transform of a signal $y[n]$: $Y(z)$.
- Vector: $\mathbf{v}$.
- Matrix: $\mathbf{A}$.
- Transpose of matrix $\mathbf{A}$: $\mathbf{A}^T$.
- Modulus: $|z|^2 = zz^*$.
- Augmented vector: $\underline{\mathbf{v}}$.

As we progress through the tutorial, the notation will become familiar and clear to the reader, enhancing comprehension. This notation provides a concise and standardized framework for discussing key concepts and techniques. In the following sections, we will use this established notation to explore the theoretical foundations and practical applications of adaptive filtering for complex-valued signals.

We now turn to the next section, which focuses on the design of discrete linear filters.

II. Discrete Linear Filter Design

Filtering is a key operation in signal processing. From a mathematical modeling perspective, a filter is a function that takes an input signal and produces a processed output signal [35]–[39], and can be classified according to their characteristics as:

- Linear or non-linear.
- Discrete or continuous in time.
- Finite impulse response or infinite impulse response.
- Time variant or time invariant.
- Parametric or non-parametric.

The filters examined in this tutorial will be referred to as transversal filters [40] and will be of the linear, discrete-time, finite-response and parametric type. Their coefficients (parameters) during operation will be constant, and thus they are also time-invariant. During adjustment (while the algorithm is running), these parameters may vary in terms of the fitting function over time (parametric model). The details of the signal and filter model will be discussed in more detail later.

Processing systems generally have the output signal of the same type as the input signal (either continuous or discrete), but this is not mandatory. It is possible to construct a system where a certain type of signal results in another type of signal. For example, sampling a signal is a process that converts a continuous input signal into a discrete output signal. Conversely, it is also possible to reconstruct a signal, allowing filters where the input signal and the response signal have different natures [41].

This tutorial will focus on the examination of discrete complex signals and transversal complex filters, a specific case of a system where both the input and output signals belong to the same representation. Each of these will be described in its respective subsections below.

A. Discrete Modeling

A discrete-time signal is a representation of information that is measured only at specific time instants. A discrete signal will only have defined values at specific points along a discrete sequence of time instants, even if the values taken by the function can be continuous. Mathematically, a discrete signal is a function $u[n]$ of the form

$$u : \mathcal{U} \in \mathbb{Z} \to \mathbb{F}.$$

Here, $\mathcal{U}$ represents an interval of integers $n \in \mathbb{Z}$. $\mathbb{F}$ is the field considered and this will determine the nature of the signal. For example, a real signal is characterized by the field $\mathbb{F} = \mathbb{R}$, while a complex signal is represented by $\mathbb{F} = \mathbb{C}$. Similarly, a quaternion signal is given by $\mathbb{F} = \mathbb{H}$, and so on.

A discrete signal can be obtained through an operation called sampling, performed on a continuous signal. The sampling values are typically evenly spaced in time. Fig. 2 provides a simple illustration of the discretization process for a

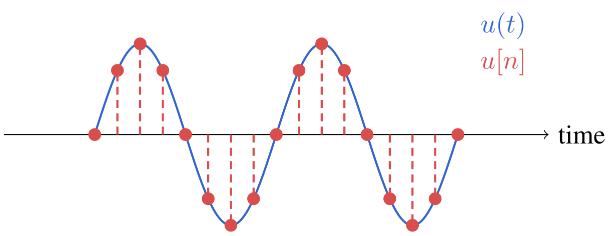


Fig. 2. An illustrative example of discretizing a sine wave $u(t)$ (blue). The red circles represents the waveform sampled in discrete time $u[n]$.

real signal, showing the nature of the discrete (red) in contrast to its continuous counterpart (blue). Discrete-time signals will be represented with brackets in their time variable, for example $u[n]$. Signals can have more than one parameter, for example, $u[n, m]$, but such cases will not be considered in the present approach. This tutorial will focus specifically on discrete one-parameter signals, particularly those represented in complex form as $u[n] = x[n] + jy[n]$, i.e., signals belonging to $\mathbb{F} = \mathbb{C}$.

So far, we have briefly discussed the nature of signals. The next step is to define the systems used to perform operations on these signals. This will be accomplished by introducing the definition of linear systems, which can be used in complex filtering.

B. Linear Modeling

From a mathematical perspective, the linear filter is the most basic filter structure in signal processing. In this category, signals are linearly combined, resulting in an expression that constitutes a linear discrete difference equation with constant coefficients [42]. The general structure of a linear filter, with a single input and a single output (abbreviated SISO), is shown in Fig. 3. It can be applied to both real and complex discrete signals, depending on the chosen coefficients $a_i, b_j \in \mathbb{F}$, where $i \in \{0, 1, \ldots, M-1\}$ and $j \in \{0, 1, \ldots, N-1\}$. If real coefficients are employed, the structure remains real, whereas the use of complex coefficients results in a complex structure. It is important to note that the complex structure employed is derived through the direct extension of the real structure. However, it is essential to note that simply using the structure as an extension of real numbers may not be advantageous in some cases, as this tutorial seeks to demonstrate. For a more in-depth analysis of complex signals, the development of new designs beyond a mere extension of the real structure becomes necessary. This development will be explored in the following sections. In this section, we will use the linear structure, which is a complex extension of the real case, as the foundation for our analyses.

The blocks labeled z^{-1} are referred to as unit delays, as they represent a delay of one discrete-time step in the Z-transform domain [36]. A unit delay block takes an input signal $u[n]$ and produces its delayed version $u[n-1]$.

The combination of inputs and outputs presented in Fig. 3 can be written as

$$v[n] = \sum_{k=0}^{M-1} a_k u[n-k] + \sum_{k=1}^{N} b_k v[n-k], \tag{1}$$

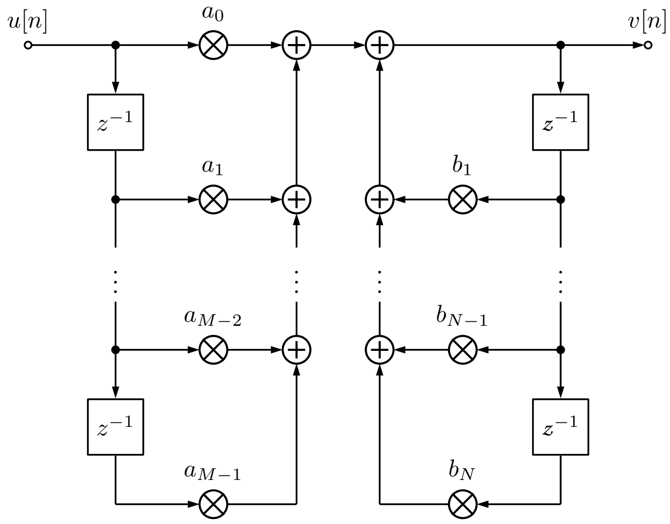


Fig. 3. Diagram of a general discrete linear filter with $M \times N$ coefficients. The input signal is represented by $u[n]$ and the output is $v[n]$.

Equation (1) is the general form of a filter with a finite number of coefficients, it is the discrete time mathematical model of a discrete linear filter. The word "linear" essentially refers to the type of finite-difference equation obtained. If the coefficients a_k and b_k, as well as the signals u (subsequently v), are all real, the filter is called a *real linear discrete filter*. If they are complex, the filter is referred to as a discrete linear complex filter.

Linear discrete filters can be categorized into two main types: filters with infinite impulse response (IIR) and filters with finite impulse response (FIR). This classification will be discussed in the following section.

C. IIR and FIR Filters

An infinite impulse response (IIR) filter depends not only on current and past input values but also on past outputs, through feedback terms scaled by coefficients b_k. This feedback structure results in an impulse response that theoretically extends infinitely over time, as described in Equation (1). The general expression, where $b_k \neq 0$ for some k, corresponds to an autoregressive moving average (ARMA) model.

In contrast, the filters employed in this tutorial are finite impulse response (FIR) filters, which constitute a special case of IIR filters where all feedback coefficients b_k are zero. FIR filters rely solely on a finite number of past inputs and do not incorporate feedback. They are also referred to as *transversal filters* and correspond to a moving average (MA) model.

Applying FIR modeling condition to Equation (1) and renaming the input constants as[1] $a_k = w_k^*$, we obtain the general expression of the transversal filter,

$$v[n] = \sum_{k=0}^{M-1} w_k^* u[n-k] \ . \tag{2}$$

[1]There is no issue in considering the coefficients as conjugates beforehand. If such consideration is not made, the solution obtained for the filter will be the conjugate of the case where such consideration was made.

Note that the expression of a transversal filter results in a finite impulse response, as it depends solely on a linear combination of current and past input samples. This characteristic ensures that the filter's output settles to zero after a finite number of time steps following the application of an impulse input. This finite impulse response ensures inherent stability and predictable behavior, making FIR filters particularly advantageous in a wide range of signal processing applications [43], [44].

The previous expression can be condensed by representing the coefficients as a column vector $\mathbf{w} = [w_0 \ \dots \ w_{M-1}]^T$ and the signal as a column vector $\mathbf{u}[n] = [u[n] \ \dots \ u[n-M+1]]^T$, both being complex and of dimension M. The expression provided in (2) can now be written as

$$v[n] = \mathbf{w}^H \mathbf{u}[n]. \tag{3}$$

The parameter M, denoting the number of coefficients, is termed the "size" or "length" of the transversal filter. A visual representation of this configuration is illustrated in Fig. 4.

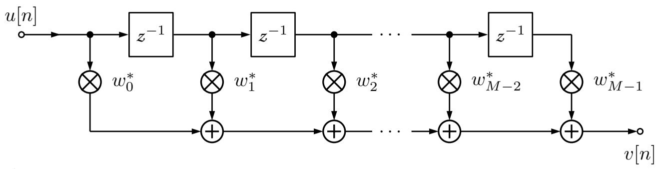


Fig. 4. Discrete FIR transversal filter of size M. This filter corresponds to the left side of Fig. 3 and does not depend on output values at any instant.

The transversal filtering described in Fig. 4 can be graphically abstracted as illustrated in Fig. 5, where the input signal $u[n]$ passes through the set of filter coefficients $\mathbf{w}$, resulting in the output signal $v[n]$. This abstraction serves as an important building block in adaptive structures designed for complex signals.

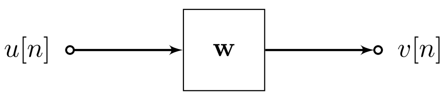


Fig. 5. Example of a signal $u[n]$ at some instant n filtered by a FIR (transversal) denoted by $\mathbf{w}$, resulting into a scalar signal $u[n]$.

Transversal filters will serve as the fundamental model throughout this tutorial.

Fig. 6 elucidates the interconnection between each category, presenting a visual guide to our exploration.

We have progressed through each category, and now we will examine the specifics of complex linear transversal filters, simply referred to as *transversal filters*. In the next section, we will explore in more detail the applications in which these filters can be employed, such as system identification, equalization, prediction, and noise cancellation.

III. Transversal Filtering Configurations

This section covers some fundamental applications where transversal filters find extensive use. Among these applications, we highlight system identification, equalization, prediction, and noise canceling. In addition, different types of

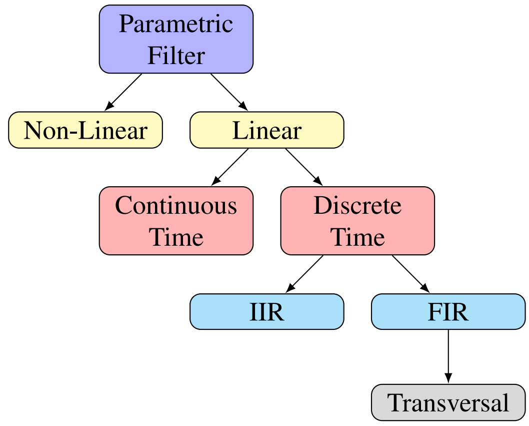


Fig. 6. Diagram for the definition of the transversal filter, valid for real or complex parametric filters. Note that in other terminations, there may be branches not shown here.

cost strategies and algorithms, which play a crucial role in these applications, will be discussed. The goal is to provide a comprehensive understanding of the diverse ways transversal filters are used, along with the evaluation criteria that guide their implementations in specific scenarios.

We will explore four basic filter configurations as described in [45]. The complex error is a scalar denoted by $e_L[n]$ (L indicates that the error originates from a linear filter configuration). This error is used by some adaptive algorithm for the recursive adjustment of the filter coefficients, which is indicated in the figures by a dashed line. The input signal of the filter is denoted by $u[n]$ and the output by $v[n]$. The "desired" (or training) signal is represented by $d[n]$, and the filter is characterized by the coefficient vector $\mathbf{w}$.

A. System identification

In system identification, the transversal filter is used to linearly approximate a system (e.g. communication channel) that one wants to estimate. Note that this approximation is a linear representation with minimal error. Fig. 7 illustrates this configuration.

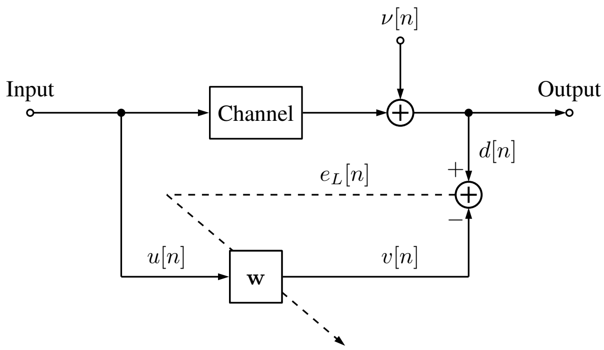


Fig. 7. Usual configuration for a channel identification. The minimization of error turns the filter (indicated as $\mathbf{w}$) a linear estimated copy of the channel.

Here, the input signal $u[n]$ fed both: the channel $h[n]$ (unknown) and the adaptive filter with coefficients $\mathbf{w}$. The channel output is $d[n] = u[n] * h[n] + \nu[n]$, where $\nu[n]$ is additive noise. The filter output is $v[n] = \mathbf{w}^T\mathbf{u}[n]$. The linear complex error $e_L[n] = d[n] - v[n]$ is the difference between the channel output and the filter output.

Identifying an unknown system allows the creation of mathematical models and parameter estimation to represent the behavior of that system.

B. Channel Equalization

The equalization process, illustrated in Fig. 8, is designed to make the linear filter act as the inverse of the specified channel. The objective is to ensure that the signal emerging from the channel closely resembles the original transmitted signal.

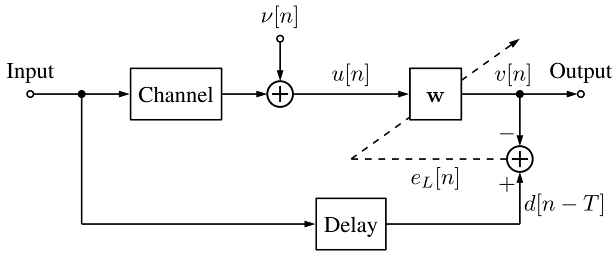


Fig. 8. Usual configuration for a channel equalization. The minimization of error turns the filter (indicated as $\mathbf{w}$) a estimated copy of the channel in some cases).

The input signal $x[n]$ passes through a channel characterized by its impulse response $h[n]$, resulting in the received signal $u[n] = x[n]*h[n]+\nu[n]$, where $\nu[n]$ represents additive noise. The filter, defined by the coefficient vector $\mathbf{w}$, processes $u[n]$ to generate the output $v[n] = \mathbf{w}^T\mathbf{u}[n]$. The error signal is then computed as $e_L[n] = d[n-T] - v[n]$, where $d[n-T]$ is a delayed version of the original input signal $x[n]$ and serves as the desired response.

In this case, finding the filter coefficients is equivalent to approximating the inverse of the system, providing a mathematical approach to computing the inverse of the convolution process through an iterative method.

Equalization is a fundamental operation that aims to ensure accurate and high-quality reception of information transmitted through a communication channel at the receiver side.

C. Prediction

In linear prediction, the goal is to forecast future signal values using past samples via an adaptive filter, typically achieved through an iterative method. The filter possesses the ability to adapt its coefficients, aiming to minimize prediction errors. Fig. 9 depicts the configuration for the predictor indicated by $\mathbf{w}$.
The input signal $d[n]$ is delayed by one sample through a unit delay z^{-1}, resulting in $u[n] = d[n-1]$. The transversal filter $\mathbf{w}$, processes $u[n]$ to generate the predicted signal $v[n] = \mathbf{w}^T\mathbf{u}[n]$. The prediction error is defined as $e_L[n] = d[n]-v[n]$, where $x[n]$ is the desired signal (current sample).

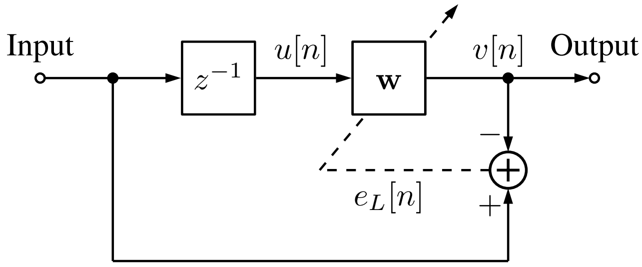


Fig. 9. Usual configuration for a signal prediction. The coefficients $\mathbf{w}$ are adjusted by minimizing the prediction error $e_L[n]$, which is determined by the signal and its delayed version.

D. Noise canceling

In noise canceling, the goal is to estimate and remove the noise that has contaminated a desired signal; see Fig. 10. Let $s[n]$ represent the desired signal. The noise is present in both inputs, but due to the different locations of the sensors, we observe two distinct input signals:

- the primary signal: $d[n] = s[n] + \nu_0[n]$,
- the reference noise signal: $u[n] = \nu_1[n]$, correlated with $\nu_0(n)$.

The error is defined as $e_L[n] = d[n] - \hat{\nu}_0[n]$, where $\hat{\nu}_0[n] = \mathbf{w}^H\nu_1[n]$ is an estimate of the noise derived from $\nu_1[n]$, the output approximates the clean signal $s[n]$ with the noise effectively suppressed.

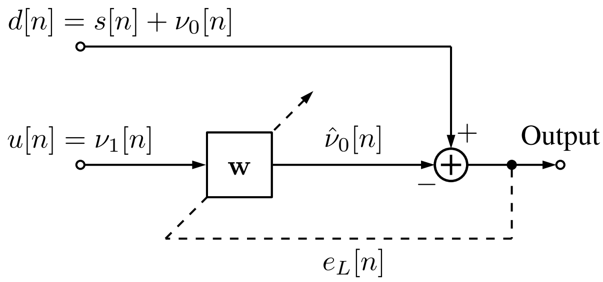


Fig. 10. Usual configuration for noise canceling.

The examples presented so far illustrate the versatility of transversal linear filters, demonstrating various applications that employ the same coefficient structure represented by $\mathbf{w}$. In the following, we discuss different types of cost function, namely scalar functions of the error, which provide a criterion for adjusting the filter coefficients during the adaptation process.

E. Error Minimization and Cost Functions

In adaptive filtering processes, either using iterative algorithms or direct solutions, coefficient adjustment is typically carried out through error $e[n]$, which is used to construct a function to be minimized (or sometimes maximized). This function is known as the cost function (or objective function), denoted here by the letter $\mathcal{J}$. The search for critical points of this function leads to optimal solution parameters (local at least and, in the case of quadratic expressions, global). Some commonly used examples of such functions include:

- CMSE - Complex Mean Squared Error. It uses the absolute mean square error at each time step as the function to be minimized. Its an extension of MSE to complex cases,
$$\mathcal{J}_{\text{CMSE}} = \text{E}[|e[n]|^2], \tag{4}$$
where $|e[n]|^2 = e[n]e_L^*[n]$.
- CNMSE - Complex Normalized Mean Squared Error. Its normalizes the above CMSE by the energy of desired signal,
$$\mathcal{J}_{\text{CNMSE}} = \frac{E[|e[n]|^2]}{[|d[n]|^2]}.$$
- CMAE - Complex Mean Absolute Error. Mean of absolute samples of error at a given instant,
$$\mathcal{J}_{\text{CMAE}} = E[|e[n]|].$$
- Gaussian Entropy Criterion - Described in [46], defined as
$$\mathcal{J}_{\text{ENT}} = \left[\text{E}\{|e[n]|^2\}\right]^2 - |\text{E}\{e^2[n]\}|^2.$$

These are just a few examples of cost functions. Here we conclude the exploration of the basic knowledge of cost functions. Understanding these functions is crucial for optimizing the performance of filters in various applications.

The next natural step involves employing these cost functions to directly, or by iterative methods, calculate the filter coefficients, as previously mentioned. To achieve this, it is essential to dedicate time to studying a few of complex calculus. The effort invested in understanding these definitions upfront will save time during subsequent calculations, allowing us to conduct all computations in the complex domain. In the upcoming section, we will explore the realm of complex calculus, an indispensable tool for advanced analyses in complex adaptive filtering.

IV. Wirtinger Calculus

In the preceding section, we discussed various cost functions that lead to the minimization of the error. Now, we investigate the Wirtinger calculus, also known as the $\mathbb{CR}$ calculus, which serves as an exceptional tool for calculating derivatives of expressions that depend on complex numbers. This calculus allows for manipulation and optimization without the need for a tedious expansion to real components.

In all the filtering schemes illustrated earlier, the error is defined as $e[n] = d[n]-v[n]$, where the filter output is given by $v[n] = \mathbf{w}^H\mathbf{u}[n]$. Consequently, the cost function (as described in Section III-E) become a function of the filter coefficients, that is, $\mathcal{J} = \mathcal{J}(\mathbf{w})$. This function is real-valued (minimization requires an ordered field) with a complex variable $\mathbf{w}$. Since this type of function is not holomorphic [12], [47], it requires special treatment provided by the Wirtinger calculus [48].

Toward the end of this section, we apply the Wirtinger calculus to calculate the optimal solution for the transversal filter, providing a practical example of its utility.

A. Basics of Wirtinger Calculus

The Wirtinger calculus is useful for calculating derivatives of functions with complex variables, streamlining the notation, and circumventing the need to rewrite all calculations in terms of their real and imaginary components. The name pays homage to the German mathematician Wilhelm Wirtinger, who developed this technique [48].

Let f be a complex function with variable $w = x + jy$, where $x, y \in \mathbb{R}$, whether holomorphic or not. The fundamental concepts of Wirtinger calculus can be summarized in two points:

- Express functions, whether real or complex, in terms of the complex variable w along with its conjugate w^*. Treat both the variable and its conjugate as independent variables and write the function f as $f(w, w^*)$. We refer to the augmented variable as $\underline{w} = (w, w^*)$, using an underline below the corresponding letter to denote such pairs. This new variable $\underline{w}$ belongs to the space denoted as $\mathbb{C}^2_*$, which represents a complexified isomorphism of $\mathbb{R}^2$ (or equivalently $\mathbb{C}$, since complex numbers are also isomorphic to $\mathbb{R}^2$). Here, $\mathbb{C}^2_*$ is a subspace of $\mathbb{C}^2$, consisting of pairs (w, w^*) where w^* is the complex conjugate of w. Unlike a general element of $\mathbb{C}^2$, where both components can be arbitrary complex numbers, elements of $\mathbb{C}^2_*$ are constrained by this conjugation relationship, effectively reducing degrees of freedom to match those of $\mathbb{R}^2$.
- Facilitate the calculation of derivatives for functions expressed as described in the previous item by applying the standard rules of calculus. This is achieved with the aid of the operators defined as
$$\frac{\partial}{\partial w} \triangleq \frac{1}{2}\left(\frac{\partial}{\partial x} - j\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial w^*} \triangleq \frac{1}{2}\left(\frac{\partial}{\partial x} + j\frac{\partial}{\partial y}\right). \tag{5}$$
Intuitively, the derivative $\partial f/\partial w$ measures the sensitivity of f with respect to the "holomorphic direction" of w, while $\partial f/\partial w^*$ captures the sensitivity in the "antiholomorphic direction". Hence, the presence of derivatives with respect to w^* indicates that the function is not holomorphic, but still differentiable in the Wirtinger sense. These operators enable the computation of the gradient, divergence, Laplacian, and other mathematical operators involving complex numbers directly in complex notation, without requiring explicit expressions in terms of real components [16].

Wirtinger calculus is particularly essential in complex-valued signal processing because it eliminates the need to decompose complex variables into their real and imaginary parts, which can complicate derivative calculations. In applications like the optimization of complex transversal filters, this approach not only simplifies the mathematical formulation but also reduces computational overhead, making it indispensable for efficient algorithm design in fields such as wireless communications and adaptive filtering [29], [49]. Wirtinger calculus is also very relevant for operations on holomorphic and non-holomorphic functions, especially for the latter case in which the functions of the real components have no relation to those of the imaginary components of the complex function, thus not satisfying the Cauchy–Riemann conditions [16], [48]. This is the case for complex functions that have only the real or imaginary part.

As a illustrative example, let's consider the function $f(w) = |w|^2$. When expressed in terms of w and w^*, we have $f(w, w^*) = ww^*$. This function can also be represented as $f(x, y) = x^2 + y^2$ through the change of variables: $x = (1/2)(w + w^*)$ and $y = (j/2)(w^* - w)$. Calculating the derivative of f with respect to the complex variable w by treating it as a usual independent variable, we obtain

$$\frac{\partial f(w, w^*)}{\partial w} = \frac{\partial}{\partial w}(ww^*) = w^* \ .$$

Similarly, with respect to the conjugate component w^*,

$$\frac{\partial f(w, w^*)}{\partial w^*} = w \ .$$

These calculations can also be performed using the right-hand side of the expressions in (5), which produces the same results. It is important to note that, although w and w^* are not truly independent variables (since w^* is the complex conjugate of w), Wirtinger calculus treats them as formally independent from a functional standpoint. This is valid as long as the function is differentiable as a function of two real variables x and y, thus enabling more flexible and algebraic manipulation.

The relationship between the variables in either their real or augmented forms is described by a complex transformation matrix,

$$\begin{bmatrix} w \\ w^* \end{bmatrix} = \begin{bmatrix} x + jy \\ x - jy \end{bmatrix} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} . \tag{6}$$

The matrix present in the above expression is denoted as

$$\mathbf{T}_s = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} , \tag{7}$$

represents the transformation $\mathbb{R}^2 \to \mathbb{C}^2_*$ that maps the canonical vectors $\mathbf{e}_1 = [1 \quad 0]^T$ and $\mathbf{e}_2 = [0 \quad 1]^T$ to the complex vectors $[1 \quad 1]^T$ and $[j \quad -j]^T$, respectively.

The transformation matrix $\mathbf{T}_s$ presented in Equation (7) represents the central role in mapping real coordinates $(x, y) \in \mathbb{R}^2$ to the augmented complex variable $\underline{w} = (w, w^*) \in \mathbb{C}^2_*$. This matrix is invertible, with inverse given by $\mathbf{T}_s^{-1} = \frac{1}{2}\mathbf{T}_s^H$, since $\mathbf{T}_s\mathbf{T}_s^H = 2\mathbf{I}$.

Geometrically, $\mathbf{T}_s$ defines a linear transformation that aligns the real plane with the complex domain under the conjugation constraint, thus enabling the use of Wirtinger calculus, which treats w and w^* as independent variables. This is critical when computing gradients of real-valued functions defined on complex variables. A more detailed mathematical analysis on this topic can be found in [28], [50].

B. Extension to the Multivariate Complex Case

The principles of Wirtinger calculus can be extended to functions of multiple complex variables. Consider a complex vector $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_N \end{bmatrix}^T \in \mathbb{C}^N$, where each component is expressed as $w_k = x_k + jy_k$, with $x_k, y_k \in \mathbb{R}$ for all $k \in \{1, 2, \dots, N\}$. The augmented complex vector can be constructed by stacking the complex vector $\mathbf{w}$ and its conjugate as

$$\underline{\mathbf{w}} \triangleq \begin{bmatrix} \mathbf{w} \\ \mathbf{w}^* \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ w_1^* \\ \vdots \\ w_N^* \end{bmatrix}_{2N \times 1} . \tag{8}$$

It is also common to employ a notation in which the vector is formed by alternating values with their complex conjugates, denoted $\mathring{\mathbf{w}} = \begin{bmatrix} w_1 & w_1^* & \cdots & w_N & w_N^* \end{bmatrix}^T$. This notation, as seen in works such as [51], [52], provides an alternative representation. Similarly, a linear transformation can relate $\underline{\mathbf{w}}$ to $\mathring{\mathbf{w}}$, however, for consistency and clarity, this tutorial will adhere to the notation presented in equation (8).

Similarly to the case described in (6), there exists a complex matrix that relates the augmented vector to the real one by

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{w}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} . \tag{9}$$

Here, $\mathbf{I}$ represents the identity matrix of dimension $N \times N$, and the complex vector is denoted as $\mathbf{w} = \mathbf{x} + j\mathbf{y}$, with $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_N \end{bmatrix}^T$, both in $\mathbb{R}^N$. Equation (9) can be expressed more concisely in terms of the augmented variable $\underline{\mathbf{w}}$ and the real composite vector $\mathbf{r} = \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix}^T$ as

$$\underline{\mathbf{w}} = \mathbf{T}\,\mathbf{r}, \tag{10}$$

where the complex transformation matrix is

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} , \tag{11}$$

with the inverse given by $\mathbf{T}^{-1} = \frac{1}{2}\mathbf{T}^H$. The matrix $\mathbf{T}$ relates the linear space $\mathbb{R}^{2N}$ to the space denoted by $\mathbb{C}^{2N}_*$. Fig. 11 illustrates the transformation visually.

Remember that, in this representation, we have $\mathbb{C}^{2N}_* \subset \mathbb{C}^{2N}$. The augmented space $\mathbb{C}^{2N}_*$ can be decomposed as

$$\mathbb{C}^{2N}_* = \mathcal{W}^{1,0} \oplus \mathcal{W}^{0,1},$$

where the symbol $\oplus$ denotes the direct sum of the subspaces $\mathcal{W}^{1,0}$ and $\mathcal{W}^{0,1}$ [53]. The complex-linear subspace $\mathcal{W}^{1,0}$ is defined as

$$\mathcal{W}^{1,0} = \left\{ \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} : \mathbf{w} \in \mathbb{C}^N \right\},$$

and the subspace $\mathcal{W}^{0,1}$, consisting of conjugate-linear (or anti-linear) components, is given by

$$\mathcal{W}^{0,1} = \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{w}^* \end{bmatrix} : \mathbf{w} \in \mathbb{C}^N \right\}.$$

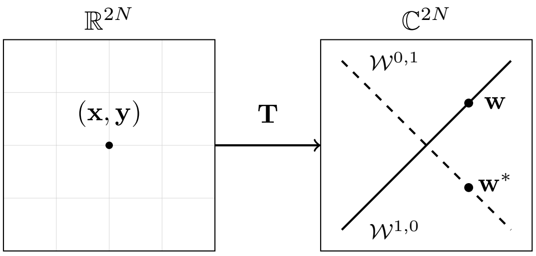


Fig. 11. Illustration of the mapping from the real space $\mathbb{R}^{2N}$ to the augmented complex space $\mathbb{C}_*^{2N} \subset \mathbb{C}^{2N}$. The real plane, represented by the rectangle on the left, is mapped into the direct sum of the linear subspace $\mathcal{W}^{1,0}$ (solid line) and the conjugate-linear subspace $\mathcal{W}^{0,1}$ (dashed line) on the right. These subspaces span the image of the real data in the augmented space, i.e., in $\mathbb{C}_*^{2N} = \mathcal{W}^{1,0} \oplus \mathcal{W}^{0,1}$. Note that although the subspaces $W^{1,0}$ and $W^{0,1}$ are represented as lines in the figure, they are complex spaces, and the depiction is schematic.

Complex augmented vectors play a significant role in statistical modeling and in the design of augmented filters, which are capable of capturing the pseudo-correlation of the received signal, as will be further explored in subsequent sections.

One can define the column operators as

$$\frac{\partial f}{\partial \mathbf{w}^T} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_N} \end{bmatrix} \quad \text{and} \quad \frac{\partial f}{\partial \mathbf{w}^H} = \begin{bmatrix} \frac{\partial f}{\partial w_1^*} \\ \vdots \\ \frac{\partial f}{\partial w_N^*} \end{bmatrix}, \tag{12}$$

where each element is a Wirtinger operator as given in (5). It is also possible to define the row versions of (12) as

$$\frac{\partial f}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial f}{\partial w_1} & \cdots & \frac{\partial f}{\partial w_N} \end{bmatrix} \quad \text{and} \quad \frac{\partial f}{\partial \mathbf{w}^*} = \begin{bmatrix} \frac{\partial f}{\partial w_1^*} & \cdots & \frac{\partial f}{\partial w_N^*} \end{bmatrix}. \tag{13}$$

The definitions presented in (12) and (13) are preferable because they allow for straightforward algebraic manipulation of complex vector expressions. In this approach, the complex variable appearing in the denominator is differentiated in the function on which the operator acts. This permits treating $\mathbf{w}$, $\mathbf{w}^*$, $\mathbf{w}^T$, and $\mathbf{w}^H$ as independent variables, enabling direct application of differentiation rules and simplifying symbolic manipulations in complex-valued functions.

As a simple example to illustrate the use of vector derivatives, consider the function

$$f = \mathbf{w}^H \mathbf{R} \mathbf{w}, \tag{14}$$

a quadratic form commonly encountered in optimization problems. Let $\mathbf{w} \in \mathbb{C}^M$ and $\mathbf{R} \in \mathbb{C}^{N \times N}$. We can write

$$\frac{\partial f}{\partial \mathbf{w}^H} = \frac{\partial}{\partial \mathbf{w}^H} \left(\mathbf{w}^H \mathbf{R} \mathbf{w} \right),$$

treating $\mathbf{w}^H$ as an independent variable with respect to $\mathbf{w}$. In this case, we apply the standard differentiation rules for monomials, which yields

$$\frac{\partial f}{\partial \mathbf{w}^H} = \mathbf{R}\mathbf{w}.$$

Note that $\partial f / \partial \mathbf{w}^H$ is constructed from the components $\partial f / \partial w_k^*$, for $k = 1, \ldots, N$, arranged to form a column vector. This ensures that the result of the derivative operation is a column vector, which is consistent with the expected outcome.

Using the same function given in (14), we can compute the derivative with respect to the holomorphic component as

$$\frac{\partial f}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{w}^H \mathbf{R} \mathbf{w} \right) = \mathbf{w}^H \mathbf{R},$$

which results in a row vector, consistent with the standard differentiation rules. Alternatively, to compute the derivative $\partial f / \partial \mathbf{w}^T$, we consider the transposed version of the quadratic form, which remains unchanged due to its invariance under transposition. In fact, we have

$$f = \mathbf{w}^H \mathbf{R} \mathbf{w} = \left(\mathbf{w}^H \mathbf{R} \mathbf{w} \right)^T = \mathbf{w}^T \mathbf{R}^T \mathbf{w}^*,$$

and thus the derivative is

$$\frac{\partial f}{\partial \mathbf{w}^T} = \mathbf{R}^T \mathbf{w}^*.$$

The expression represents the column vector associated with the derivative. Similarly, using the form in (14) for the remaining case, we have

$$\frac{\partial f}{\partial \mathbf{w}^*} = \frac{\partial}{\partial \mathbf{w}^*} \left(\mathbf{w}^T \mathbf{R} \mathbf{w}^* \right) = \mathbf{w}^T \mathbf{R}^T.$$

This result can also be verified by performing the derivative component-wise. Moreover, the row derivatives could have been obtained from their corresponding column versions (and vice versa), since the expressions in (12) and (13) satisfy the relationships

$$\left(\frac{\partial f}{\partial \mathbf{w}^T} \right)^T = \frac{\partial f}{\partial \mathbf{w}} \quad \text{and} \quad \left(\frac{\partial f}{\partial \mathbf{w}^H} \right)^T = \frac{\partial f}{\partial \mathbf{w}^*}, \tag{15}$$

by construction. It is important to note that the previous equalities (15) hold for transposition, which is valid for any scalar complex function.

More generally, when taking the Hermitian these equalities only hold if $f = f^*$, i.e. $f \in \mathbb{R}$, in this case we can write

$$\left(\frac{\partial f}{\partial \mathbf{w}} \right)^H = \frac{\partial f}{\partial \mathbf{w}^H} \quad \text{and} \quad \left(\frac{\partial f}{\partial \mathbf{w}^*} \right)^H = \frac{\partial f}{\partial \mathbf{w}^T}.$$

In the next section, we introduce the derivatives in their augmented form, which make use of the vector operators discussed so far.

C. Augmented Expressions

Building upon this approach, one can define another pair of derivatives with respect to the augmented variable by placing each row derivative (13) side by side as

$$\frac{\partial f}{\partial \underline{\mathbf{w}}} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{w}} & \frac{\partial f}{\partial \mathbf{w}^*} \end{bmatrix} \tag{16}$$

or defining the Hermitian version by making a new column of operators from (12),

$$\frac{\partial f}{\partial \underline{\mathbf{w}}^H} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{w}^H} \\ \frac{\partial f}{\partial \mathbf{w}^T} \end{bmatrix}. \tag{17}$$

This approach extends naturally to the real composite case, where the derivatives of a function $f(\mathbf{r})$ can be expressed as

$$\begin{aligned}\frac{\partial f}{\partial \mathbf{r}} &= \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}} & \frac{\partial f}{\partial \mathbf{y}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_N} & \frac{\partial f}{\partial y_1} & \cdots & \frac{\partial f}{\partial y_N} \end{bmatrix}\end{aligned} \tag{18}$$

and

$$\frac{\partial f}{\partial \mathbf{r}^T} = \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}^T} \\ \frac{\partial f}{\partial \mathbf{y}^T} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \\ \frac{\partial f}{\partial y_1} \\ \vdots \\ \frac{\partial f}{\partial y_N} \end{bmatrix}. \tag{19}$$

Using the matrix defined in (11), it is possible to relate equations (17) and (19) through the following identity:

$$\frac{\partial f}{\partial \mathbf{r}^T} = \mathbf{T}^H \frac{\partial f}{\partial \underline{\mathbf{w}}^H}. \tag{20}$$

D. Second order Expressions

The adopted notation can also be extended to second-order operators. An important example is the Hessian, which can be represented using the operators defined in the previous section. Consider (19) as a column vector operator acting on (18). This allows the Hessian of a real function f to be written as

$$\begin{aligned}\mathbf{H} &= \frac{\partial}{\partial \mathbf{r}^T} \cdot \frac{\partial f}{\partial \mathbf{r}} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^T} \\ \frac{\partial}{\partial \mathbf{y}^T} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}} & \frac{\partial f}{\partial \mathbf{y}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial f}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial f}{\partial \mathbf{y}} \\ \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial f}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}^T} \cdot \frac{\partial f}{\partial \mathbf{y}} \end{bmatrix},\end{aligned} \tag{21}$$

at some point where second-order derivatives exist. Each element of the matrix (21) can be evaluated as follows:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}^T} \cdot \frac{\partial f}{\partial \mathbf{y}} &= \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_N} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial y_1} & \cdots & \frac{\partial f}{\partial y_N} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial y_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial y_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial y_1} & \cdots & \frac{\partial^2 f}{\partial x_N \partial y_N} \end{bmatrix}.\end{aligned} \tag{22}$$

The matrix of second-order derivatives can also be constructed using the complex operators (17) and (16) with respect to a real function of complex variable as

$$\frac{\partial}{\partial \underline{\mathbf{w}}^H} \cdot \frac{\partial f}{\partial \underline{\mathbf{w}}} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{w}^H} \\ \frac{\partial}{\partial \mathbf{w}^T} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial \mathbf{w}} & \frac{\partial f}{\partial \mathbf{w}^*} \end{bmatrix}. \tag{23}$$

Equations 22 and 23 can be correlated using the matrix $\mathbf{T}$, provided that the function is real. This form is related to the Hessian in its augmented complex form, as will be seen later.

E. Example: Applying Complex Augmented Notation

Consider the real quadratic scalar function $f(\mathbf{w})$ with complex vector variable $\mathbf{w}$ given by

$$f(\mathbf{w}) = c + \mathbf{w}^H\mathbf{a} + \mathbf{a}^H\mathbf{w} + \mathbf{w}^H\mathbf{C}\mathbf{w}, \tag{24}$$

for $c \in \mathbb{R}$, a constant vector $\mathbf{a} \in \mathbb{C}$ and $\mathbf{C}$ is a Hermitian matrix, i.e. $\mathbf{C} = \mathbf{C}^H$. This function frequently arises in the context of minimizing the mean-square error for complex filters. Now, we will derive the column derivative of f, given by (24), with respect to the conjugated variable $\mathbf{w}^H$,

$$\frac{\partial f}{\partial \mathbf{w}^H} = \mathbf{a} + \mathbf{C}\mathbf{w},$$

This result is a column vector, obtained by treating the variable to be differentiated, $\mathbf{w}^H$, as independent of $\mathbf{w}$ while using the usual calculus rules.

Similarly, the derivative of (24) with respect to the conjugate variable $\mathbf{w}^*$ can be calculated. However, the expression must be rewritten in an equivalent form in terms of $\mathbf{w}^*$ as

$$f(\mathbf{w}) = c + \mathbf{a}^T\mathbf{w}^* + \mathbf{a}^H\mathbf{w} + \mathbf{w}^T\mathbf{C}^T\mathbf{w}^*.$$

Taking the derivative on both sides with respect to $\mathbf{w}^*$ in the previous equation,

$$\frac{\partial f}{\partial \mathbf{w}^*} = \mathbf{a}^T + \mathbf{w}^T\mathbf{C}^T \ .$$

This result now represents a row vector (the sum of the row vectors $\mathbf{a}^T$ and $\mathbf{w}^T\mathbf{C}^T$). As mentioned before, this notation is quite practical, because the variable to be differentiated is explicitly written in the operator. Moreover, it consistently follows the same rule in both its column vector and row vector forms.

Continuing our exploration of the function f presented in (24), we can compute second-order derivatives using the same procedure as before. For instance, we will find the expression:

$$\frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{w}}.$$

Starting with the first derivative

$$\frac{\partial f}{\partial \mathbf{w}} = \mathbf{a}^H + \mathbf{w}^H\mathbf{C},$$

which is a row vector expression. The next derivative with respect to the Hermitian variable, that is, a column operator, operates as a matrix multiplication with the last expression resulting in a matrix form given

$$\frac{\partial}{\partial \mathbf{w}^H} \cdot \frac{\partial f}{\partial \mathbf{w}} = \mathbf{C}.$$

This is consistent with the expected dimension of the operation performed.

The table I below summarizes the use of each operator in relation to the given linear and quadratic forms. The left column presents the values of the function f. The upper row indicates the operators to be applied to these values.

After some development regarding Wirtinger calculus, it is possible to apply it in the optimization of complex transversal filters. The next topic explores the derivation of the optimal (Wiener) filter expression obtained by minimizing the mean square error.

Table I: Table for some applications for vector Wirtinger operators.

f	$\frac{\partial}{\partial \mathbf{w}}$	$\frac{\partial}{\partial \mathbf{w}^*}$	$\frac{\partial}{\partial \mathbf{w}^T}$	$\frac{\partial}{\partial \mathbf{w}^H}$
$\mathbf{a}^H\mathbf{w}$	$\mathbf{a}^H$	0	$\mathbf{a}^*$	0
$\mathbf{w}^H\mathbf{a}$	0	$\mathbf{a}^T$	0	$\mathbf{a}$
$\mathbf{w}^H\mathbf{C}\mathbf{w}$	$\mathbf{w}^H\mathbf{C}$	$\mathbf{w}^T\mathbf{C}^T$	$\mathbf{C}^T\mathbf{w}^*$	$\mathbf{C}\mathbf{w}$

F. The Wiener-Hopf solution

Armed with the fundamental principles of Wirtinger calculus, we can now unveil the optimal solution for transversal filters utilizing complex parameters. The calculations closely mirror those in the real domain, eliminating the need for separate treatment of real and imaginary components. Exploiting this simplification, we proceed to derive the expression for the optimal filter based on the mean square error cost function.

Utilizing the transversal filter depicted in Fig. 4, where $u[n]$ represents the input signal, $\mathbf{w}$ denotes the filter coefficients, and $v[n]$ is the output signal, we can express the error as

$$e[n] = d[n] - v[n]. \tag{25}$$

This error expression serves as the basis for minimization in any of the configurations presented. Using the mean squared error, described in Equation(4), as the cost function,

$$\mathcal{J}_{\text{CMSE}} = \text{E}\{e[n]e^*[n]\} = \text{E}\{(d[n] - v[n])(d^*[n] - v^*[n])\} \ .$$

During the adjustment, the filter coefficients are variables, and the expression for the output is a function of the filter coefficients, $v[n] = \mathbf{w}^H\mathbf{u}[n]$.

A common practice is to assume that the input and desired signal have a zero mean. This assumption facilitates some calculations. Due to this, we will make this assumption from now on. Substituting this into the earlier cost expression, we can derive the error expression in terms of the coefficients,

$$\begin{aligned}\mathcal{J}_{\text{CMSE}}(\mathbf{w}) = {} & \text{E}\{|d[n]|^2\} - \text{E}\{d[n]\mathbf{u}^H[n]\}\mathbf{w} \\ & - \mathbf{w}^H\text{E}\{d^*[n]\mathbf{u}[n]\} + \mathbf{w}^H\text{E}\{\mathbf{u}[n]\mathbf{u}^H[n]\}\mathbf{w}.\end{aligned} \tag{26}$$

To simplify the notation, we can rename the expressions of expectations as follows: the scalar variance of the desired signal (assumed to have zero mean) as $\text{E}\{|d[n]|^2\} = \sigma_d^2$, the cross-covariance vector between the desired signal and the received signal as

$$\mathbf{p}[n] = \text{E}\{d^*[n]\mathbf{u}[n]\} = \begin{bmatrix} \text{E}\{d^*[n]u[n]\} \\ \vdots \\ \text{E}\{d^*[n]u[n-M+1]\} \end{bmatrix}_{M\times 1},$$

and the covariance matrix of the input signal as

$$\begin{aligned}\mathbf{C}[n] &= \text{E}\{\mathbf{u}[n]\mathbf{u}^H[n]\} \\ &= \begin{bmatrix} \text{E}\{u[n]u^*[n]\} & \cdots & \text{E}\{u[n]u^*[n-M+1]\} \\ \vdots & \ddots & \vdots \\ \text{E}\{u[n-M+1]u^*[n]\} & \cdots & \text{E}\{u[n-M+1]u^*[n-M+1]\} \end{bmatrix}.\end{aligned} \tag{27}$$

The covariance matrix is a $M \times M$ Hermitian matrix that is positive semidefinite [45]. We will use the term "covariance" for these second-order measures not normalized by standard deviations [54]. We will also avoid the use of the term "cross" as a prefix if it is clear that it refers to the covariance between two different signals (as in the case of $\mathbf{p}$). Thus, expression (26) is composed of monomials of covariances: one scalar of the desired signal, two involving the desired signal and the input signal, and also one that employs a matrix form of the input signal.

The expression for the cost function (26) now can be written in its simplified form,

$$\begin{aligned}\mathcal{J}_{\text{CMSE}}(\mathbf{w}) = {} & \sigma_d^2[n] - \mathbf{p}^H[n]\mathbf{w} - \mathbf{w}^H\mathbf{p}[n] \\ & + \mathbf{w}^H\mathbf{C}[n]\mathbf{w}.\end{aligned} \tag{28}$$

This equation represents a quadratic form, given by the positive semidefinite Hermitian matrix $\mathbf{C}[n]$, allowing the calculation of a global minimum for coefficients $\mathbf{w}$. The statistics in $\sigma_d^2[n]$, $\mathbf{p}[n]$, and $\mathbf{C}[n]$ are characteristics of the training signal and the received signal, which makes $\mathbf{w}$ the variable that can be adjusted. Using the expression for the Hermitian derivative, Equation (12), in the cost function given by Equation (28),

$$\frac{\partial \mathcal{J}_{\text{CMSE}}}{\partial \mathbf{w}^H}(\mathbf{w}) = -\mathbf{p}[n] + \mathbf{C}[n]\mathbf{w} \ . \tag{29}$$

Assuming the ergodicity of the training signal [55], for desired signal and for noise, the statistics remain constant, it is possible to write $\mathbf{C}[n] = \mathbf{C}$ and $\mathbf{p}[n] = \mathbf{p}$ for any instant n. There is no issue in considering the training signal in this way since its design is determined by the designer.

Finding the optimal point (the point where the derivative vanishes) for the expression (29) we must have the optimal coefficients given by

$$\mathbf{w}^{\text{opt}} = \mathbf{C}^{-1}\mathbf{p}. \tag{30}$$

This expression is known as the Wiener-Hopf equation, named after the mathematicians Norbert Wiener and Eberhard Hopf, who made significant contributions to the development of filtering theory and the solution of optimal filtering problems [56].

The optimal solution has a problem which is the presence of the inverse of the covariance matrix of the filter's input signal. Iterative techniques are preferable and extensively use the gradient expression given in (29). Another issue when considering complex signals is the information loss when only considering the covariance matrix in solution. This is due to the transversal filter scheme used, which does not capture other important information that a complex signal may possess. The next section examines these details, showing that complex signals should have their properties defined in a more general manner, not just by the covariance matrix.

V. Second Order Statistics

When dealing with complex signals, it is crucial to consider certain aspects of their second-order statistics [57]. This is particularly relevant because Gaussian distributions are frequently employed to characterize signals, particularly in the context of noise modeling [58]. In this section, we explore in greater depth second-order statistics and their implications in complex signal analysis.

Considering a complex zero mean scalar signal $u[n]$ assumed to be ergodic, its scalar covariance with a delay of m samples is expressed as

$$r_u(m) = \mathrm{E}\{u[n]u^*[n-m]\} \ . \quad (31)$$

Along with this definition, the covariance matrix for the regressive vector of M samples of (31) can be represented by

$$\mathbf{C} = \begin{bmatrix} r_u(0) & r_u(1) & \dots & r_u(M-1) \\ r_u^*(1) & r_u(0) & \dots & r_u(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_u^*(M-1) & r_u^*(M-2) & \dots & r_u(0) \end{bmatrix}, \quad (32)$$

which is equivalent to the matrix described in Equation (27).

For a complex signal represented by $u[n] = x[n] + jy[n]$, the expression for the covariance at $m = 0$ in terms of its components is

$$r_u(0) = \mathrm{E}\{x^2[n]\} + \mathrm{E}\{y^2[n]\}.$$

We can define $r_x(m) = \mathrm{E}\{x[n]x[n-m]\}$ and $r_y(m) = \mathrm{E}\{y[n]y[n-m]\}$. Using these definitions, we obtain an expression that relates the signal's correlation to its correlation components at $m = 0$,

$$r_u(0) = r_x(0) + r_y(0).$$

Certainly, if the components $r_x(0)$ and $r_y(0)$ are known, it is possible to obtain the covariance $r_u(0)$ for the complex signal $u[n]$. However, the reverse is not feasible; $r_x(0)$ and $r_y(0)$ cannot be deduced solely from $r_u(0)$. Some additional information is necessary, even for any given value of m. The missing piece that connects real and complex representations is the concept of *pseudo-covariance*, sometimes referred to as complementary covariance [59] or the relation matrix element [24], defined as

$$\tau_u(m) = \mathrm{E}\{u[n]u[n-m]\}. \quad (33)$$

The previous expression is similar to the definition (31), but the conjugate is not taken here. Evaluating equation (33) for $m = 0$ one should obtain

$$\tau_u(0) = r_x(0) - r_y(0) + 2jr_{xy}(0). \quad (34)$$

The value $r_{xy}(m) = \mathrm{E}\{x[n]y[n-m]\}$ is the covariance between real and imaginary components. With the expressions (31) and (34) in hand, it is now possible to relate the real and imaginary components to the complex version. It also becomes explicit that the correlation expression (31) does not take into account the covariance between the real and imaginary components: This represents an important information that is not utilized in the solutions of the optimal filter or gradient as calculated before.

Before examining a filter capable of capturing the behavior of pseudo-covariance, it is important to spend more time on the fundamental concepts and properties of complex signals, particularly their circularity and related properties. This will be addressed in the next section.

A. Circularity and properness of a complex signal

A complex discrete signal, which is the focus of this tutorial, is a discrete time sequence formed by its real and imaginary parts [35], [36], [45]. A signal is also associated with measurement uncertainties, external noise, channel distortions, etc. Therefore, a signal received by a certain device has a probability (distribution) associated with it. The signal distortions, along with the need to recover the original signal, are precisely one of the reasons that justify filtering operations.

The fact that a complex signal is a sequence with two components that can be associated with probability distributions allows for statistical study. A signal of this type may have a distribution associated with each component, but generally signals with the *same type of distribution* in each component are considered here, even if the parameters are not the same. We will also begin the study of complex notation in statistics, specifically focusing on distributions and second-order moments. With these concepts defined, it becomes more practical to characterize signals in complex notation, similar to what was done in Section IV-A with the calculus of Wirtinger.

Consider a complex random variable of the form $Z = X + jY : \Omega_X \times \Omega_Y \to \mathbb{C}$, where both $X : \Omega_X \to \mathbb{R}$ and $Y : \Omega_Y \to \mathbb{R}$ are real random variables [60]. The symbol Ω represents the event space corresponding to each variable in the subscript. It is important to note that due to the impossibility of ordering in the complex field, the probability densities p are calculated as joint variables between X and Y, that is, $p_Z(z) \triangleq p_{X,Y}(x, y)$ [28], [49], [61], [62].

Two fundamental concepts interrelated in complex signal analysis are *circularity* and *properness* [21], [22], [26]. The general descriptions of these concepts are condensed as follows.

- **Circularity** (also includes the concept of noncircularity): In simple terms, a complex variable $Z = X + jY$ is said to be circular if its probability distribution is the same as the distribution of its rotated version $e^{j\phi}Z$ for every $\phi \in [-\pi, \pi)$. This definition requires that all moments associated with the distribution are equal in the real and imaginary parts, for any moment order. Also, the cross-moments are zero. Non-circularity is the opposite of this definition.
- **Properness** (associated with the concept of impropriety): Consider only second-order statistics and allow for easy characterization of bivariate Gaussian distributions (whether scalar or vector). The properness takes into account only second-order moments, i.e. a signal is said to be proper (second-order circular) if its pseudo-correlation is zero: this makes the cross real second-order moment zero and the real second-order moments equal in each component.

For Gaussian distributions, both the concepts of circularity and properness are equivalent because these distributions are characterized only by second-order statistics and the mean [63]. As an example, consider a bivariate Gaussian distribution denoted by $\mathcal{N}(\boldsymbol{\mu} = 0, \boldsymbol{\Sigma}_{XY})$ where $\boldsymbol{\mu} = \begin{bmatrix}\mu_X = 0 & \mu_Y = 0\end{bmatrix}^T$ is the mean vector composed of the means of each component and

$$\mathbf{\Sigma}_{XY} = \begin{bmatrix} r_X & r_{XY} \\ r_{XY} & r_Y \end{bmatrix} \tag{35}$$

is the real covariance matrix associated. The probability density function that represents the distribution is given by

$$p_{XY}(x,y) = \frac{1}{\pi(\det[\mathbf{\Sigma}_{xy}])^{1/2}} \exp\left\{-\frac{1}{2}[x \quad y]\mathbf{\Sigma}_{xy}^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right\} . \tag{36}$$

By the previous definitions, a Gaussian distribution is said to be proper if $r_{XY} = 0$ and $r_X = r_Y$. One possible measure of how circular or elliptical the Gaussian distribution is, in terms of its components, is given by the real coefficient of circularity (real version, indicated by subscript r)

$$\rho_r = \frac{r_{XY}}{\sqrt{r_X}\sqrt{r_Y}}. \tag{37}$$

with $|\rho_r| \leq 1$. More precisely,

$$\rho_r = \frac{r_{XY} - \mu_X\mu_Y}{\sqrt{r_X - \mu_X^2}\sqrt{r_Y - \mu_Y^2}},$$

for a distribution that is not centred on the origin. The expression in (37) is also known as the Pearson coefficient [64]–[67], representing the relationship between the cross-component and each energy component of the distribution.

The covariance in terms of the complex random variable $Z = X + jY$, similar to the equation (31), is

$$r_Z = \mathrm{E}\{ZZ^*\} = \mathrm{E}\{|Z|^2\} = r_X + r_Y \ .$$

The pseudo-correlation, according to (34),

$$\tau_Z = \mathrm{E}\{Z^2\} = r_X - r_Y + j2r_{XY} \ .$$

Note that each value of pseudo-covariance can be a complex number, whereas covariance values are always real. It is interesting to define a new complex circularity coefficient instead of the real Pearson. Taking the definition used in [26], [57], we have the expression of the complex correlation coefficient as

$$\rho_c = \frac{\mathrm{E}\{Z^2\}}{\mathrm{E}\{|Z|^2\}} = \frac{\tau_Z}{r_Z}. \tag{38}$$

The absolute values of ρ_c are normalized, i.e. $0 \leq |\rho_c| < 1$ [46]. The complex coefficient (38) is related to the real coefficient (37) according to [26] by

$$\rho_r = \frac{\mathrm{Im}\{\rho_c\}}{\sqrt{1 - \mathrm{Re}\{\rho_c\}}} \ .$$

It is possible to construct the augmented correlation matrix directly by taking the correlation of the augmented complex random variable $Z = \begin{bmatrix} Z & Z^* \end{bmatrix}^T$ and its hermitian as

$$\underline{\mathbf{\Sigma}} = \mathrm{E}\left\{\begin{bmatrix} Z \\ Z^* \end{bmatrix} \cdot \begin{bmatrix} Z^* & Z \end{bmatrix}\right\} = \begin{bmatrix} r_Z & \tau_Z \\ \tau_Z^* & r_Z^* \end{bmatrix} \ . \tag{39}$$

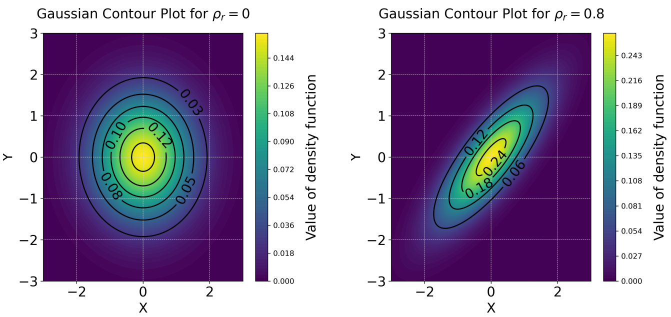
The distribution (36) in terms of augmented variables using the augmented correlation (39) is

$$p_{\underline{Z}}(\underline{z}) = \frac{1}{\pi(\det\underline{\mathbf{\Sigma}})^{1/2}} \exp\left\{-\frac{1}{2}[z^* \quad z]\underline{\mathbf{\Sigma}}^{-1}\begin{bmatrix} z \\ z^* \end{bmatrix}\right\} , \tag{40}$$

highlighting the use of complex coordinates to represent the real Gaussian distribution. In the particular case where the variable is proper, we must have $\tau_Z = 0$. This implies conditions $r_{XY} = 0$ and $r_X = r_Y$ simultaneously. The density is then expressed as in [62] by

$$p_{\underline{Z}}(\underline{z}) = \frac{1}{\pi r_z}\exp\left\{-\frac{1}{r_z}|z|^2\right\} \ .$$

The contour lines for Equation (36), equivalent to the expression in (40), are illustrated in Fig. 12. This figure features the contour plot for two Gaussian distributions with zero mean: one circular, depicted in Fig. 12(a), and another with a real circularity coefficient $\rho_r = 0.8$, presented in Fig. 12(b).



(a) (b)

Fig. 12. Contour lines for a zero-mean Gaussian with $r_z = 2$. In (a), the value of $\rho_r = 0$ indicates the circularity of the distribution, while in (b) with a value of $\rho_r = 0.8$, it shows a non-circular (elliptic) distribution.

B. *Augmented Correlation Matrix and Process*

The concepts developed so far can be used to extend the analysis to multidimensional random variables (or processes), similar to what was developed in section IV-A. This will be done here because it also introduces important notation that is used in non-circular signal analysis.

For a given random column vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} \in \mathbb{C}^N$, where each component $\mathbf{X}$ and $\mathbf{Y}$ is a real random variable with zero mean, there is a compact way to write the augmented correlation matrix, with the help of the complex augmented vector defined at (8). We can construct the random version as $\underline{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}^T & \mathbf{Z}^H \end{bmatrix}^T$ and define the complex augmented covariance matrix as

$$\begin{aligned} \underline{\mathbf{C}} = E\{\underline{\mathbf{Z}}\underline{\mathbf{Z}}^H\} &= \begin{bmatrix} E\{\mathbf{Z}\mathbf{Z}^H\} & E\{\mathbf{Z}\mathbf{Z}^T\} \\ E\{\mathbf{Z}\mathbf{Z}^T\}^* & E\{\mathbf{Z}\mathbf{Z}^H\}^* \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C} & \tilde{\mathbf{C}} \\ \tilde{\mathbf{C}}^* & \mathbf{C}^* \end{bmatrix}_{2N\times 2N} . \end{aligned} \tag{41}$$

Notice that $\mathbf{C}$ (without the underline) is the complex covariance matrix as seen before, obtained directly from the complex transversal filter design (which, instead of using a generic $\mathbf{Z}$ random vector, uses the signal as a vector). The matrix $\tilde{\mathbf{C}}$ is called the pseudo-covariance matrix. The use of the tilde (˜) above the matrix symbol is commonly employed to denote complementary matrices (pseudo-matrices concerning impropriety), and it will be used here as well. The complex augmented matrix $\underline{\mathbf{C}}$ is important in the definition of widely linear filters and can be obtained directly from its structure, as will be seen later.

The components of (41), each with dimension $N \times N$, written in terms of the real components of the complex vector, are

$$\mathbf{C} = E\{\mathbf{X}\mathbf{X}^T\} + E\{\mathbf{Y}\mathbf{Y}^T\} + j\left(E\{\mathbf{Y}\mathbf{X}^T\} - E\{\mathbf{X}\mathbf{Y}^T\}\right)$$

$$\mathbf{C} = \mathbf{C}_{XX} + \mathbf{C}_{YY} + j(\mathbf{C}_{YX} - \mathbf{C}_{XY}),$$

and

$$\tilde{\mathbf{C}} = \mathbf{C}_{XX} - \mathbf{C}_{YY} + j(\mathbf{C}_{XY} + \mathbf{C}_{YX}).$$

In the case in which the vector is proper (second-order circular) the pseudo-covariance vanishes, that is, $\tilde{\mathbf{C}} = 0$.

It is possible to characterize discrete-time Gaussian processes that vary over time through scalar covariance, which can be seen as a discrete convolution between the signal and its complex conjugate. Pseudo-covariance can be viewed in the same way without the conjugation operation. For a scalar value $Z[n]$, each element of the correlation is

$$r[k,m] = E\{Z[k+m]Z^*[k]\} - E\{Z[k+m]\}E\{Z^*[k]\},$$

and for the pseudo-correlation is

$$\tilde{r}[k,m] = E\{Z[k+m]Z[k]\} - E\{Z[k+m]\}E\{Z[k]\}\ .$$

If $\tilde{r}_z[k,m] = 0$ for every pair $[k,m]$, the process is said to be proper. If the mean is zero, the correlation coincides with the covariance of the processes. For a zero-mean WSS (Wide-Sense Stationary) process, one can write

$$r[k,m] = r[m]$$

and

$$\tilde{r}[k,m] = \tilde{r}[m],$$

depending only on the time interval between samples. For a process that can be modelled by a Gaussian WSS process, it has the property that $r[m] = r_Z$ and $\tilde{r}[m] = \tau_Z$ for all m.

Complex signals can be modeled as discrete-time processes: this is why the study of these processes is fundamental. The next topic details the signal model that will be adopted in the simulations in this tutorial, allowing for easy parameter adjustment and facilitating the understanding of the models used.

C. Signal modeling

To model and reproduce the simulations performed in this work, it is recommended to adopt a signal model defined by parameters. By adjusting these parameters, one can obtain a multitude of distributions, and for a specific set of values, you can achieve a circular (or non-circular) profile for a process (signal). This standard signal is constructed here with the general form of the adopted process in [49] with the presence of rotation factor as

$$Z[n] = e^{j\varphi}\left(\sqrt{1-\rho^2}A[n] + j\rho B[n]\right)\ . \qquad (42)$$

In this model, the complex process (signal) $Z[n]$ comprises the real components $A[n]$ and $B[n]$, both of which are scalar WSS processes with the same distribution. In the standard case, each process has a zero mean and, in general, is not correlated one with the other. The deterministic tuning parameters ρ and φ are: the profile ratio $0 < \rho < 1$, representing the weight assigned to each component and the variable $\varphi \in (-\pi, \pi]$, indicating the rotation of this profile. The signal can also be expressed as $Z[n] = X[n] + jY[n]$, where $X[n] = e^{j\varphi}\sqrt{1-\rho^2}A[n]$ and $Y[n] = e^{j\varphi}\rho B[n]$, facilitating visualization on a real/imaginary axis pair.

The relation between the profile ratio/angle present in equation (42) and the complex circularity coefficient (38) for a WSS process is given by

$$\rho_c = (1-2\rho^2)e^{j2\varphi}. \qquad (43)$$

In the case of signals $A[n]$ and $B[n]$ in (42) being uncorrelated, both being Gaussian signals with zero mean and unitary variance, this implies that the covariance matrix of a vector of M samples of $Z[n]$ is $\mathbf{C} = \mathbf{I}$, and the pseudo-covariance matrix is $\tilde{\mathbf{C}} = \rho_c\mathbf{I}$, where $\mathbf{I}$ represents the identity matrix with dimension $M \times M$. This result is similar to that found at [49], with the exception of rotation φ.

One can remember that in the case of a proper process, $\rho_c = 0$, which is equivalent to $\tau_Z = 0$. This value for ρ_c leads to a profile ratio of $\rho = 1/\sqrt{2}$ for all φ, and the components will be multiplied by the same factor and also will be uncorrelated.

D. Simulation for Signal and Noise

To illustrate the behavior of complex signals, we performed numerical simulations and computed second-order statistics. Fig. 13 and Fig. 14 display Gaussian signals and their statistics, while Fig. 15 shows the 4QAM pattern. In all cases, a set of 2×10^3 points was used and a similar analysis was employed to examine the profiles and histogram together with the covariance and pseudo-covariance figures.

The initial experiment used a circular Gaussian signal, as defined by the expression (42). For this process, both components, $A[n]$ and $B[n]$, have zero mean and unitary variance. Fig. 13(a) displays the profile of all realizations for the case where $\rho = 1/\sqrt{2}$. The signal histogram is presented in Fig. 13(c), which illustrates the frequency of values concerning the spatial distribution. The covariance values, seen as a convolution of the signal with itself, are shown in Fig. 13(b). In particular, the nearly nonexistent pseudo-covariance is observed in Fig. 13(d), as expected in the circular case. Pseudo-covariance of data is the indicator for testing the properness.

Fig. 14 depicts a similar analysis for the non-circular signal with the profile illustrated in Fig. 14(a). The histogram is presented in Fig. 14(c). In Fig. 14(b), the covariance is shown and in Fig. 14(d), the pseudo-covariance is displayed. The presence of a significant value at zero lag in the pseudo-covariance suggests that the signal is non-circular.

An example can be illustrated using a pair of 4QAM signals plus noise, as depicted in Fig. 15. Both signals in the figure

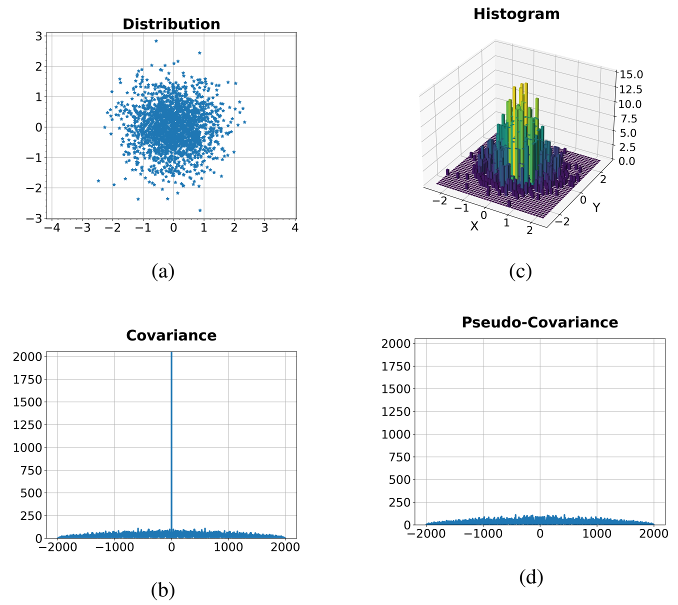


Fig. 13. Zero mean Gaussian circular signal. In (a), the profile of the signal is displayed. Figure (b) shows the corresponding histogram. (c) presents the covariance as a function of lag (horizontal axis), while panel (d) illustrates the pseudo-covariance versus lag, which is nearly zero, consistent with the properties of a circular Gaussian signal.

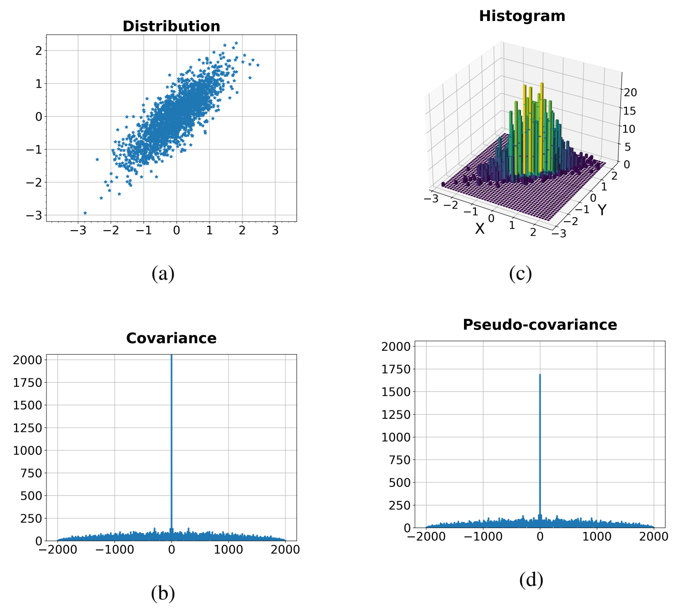


Fig. 14. Zero-mean Gaussian non-circular signal. In (a), the time-domain profile of the signal is displayed. Panel (b) shows its histogram. Panel (c) presents the covariance as a function of lag, and (d) illustrates the pseudo-covariance, also as a function of lag, which now exhibits significant values, indicative of second-order non-circularity.

exhibit circular noise modelled by the expression (42) with a scalar factor $N = 10$ as

$$Z[n] = e^{j\varphi}\left(\sqrt{1-\rho^2}A[n] + j\rho B[n]\right)/N. \tag{44}$$

Like in the last experiment, $A[n]$ and $B[n]$ have a Gaussian zero-mean distribution with unitary variance.

Fig. 15(a) shows the profile of the first 4QAM simulation. Its covariance is shown in 15(b) and its pseudo-covariance in 15(c). As expected, the symmetric 4QAM signal is proper, as can be seen in the pseudo-correlation graph. In the bottom row of the figure, the same is done for the second 4QAM realization, now with $\rho = 0.3$ and $\varphi = 0$, which generates an improper signal with profile 15(d), indicated by the pseudo-correlation shown in 15(f).

Simulations are crucial because they illustrate the characterization of signals through their second-order statistics. The presence of a non-residual value (energy) for pseudo-covariance, as depicted in Fig. 14(d) and Fig. 15(f), indicates the existence of information (the variance between real and imaginary parts) that can be considered in the processing of the considered signal.

The filter depicted in Fig. 4 is unable to capture the pseudo-covariance using the mean squared error (MSE) cost function, which can pose a problem as some information may be lost. The next section presents a filter model that can capture the pseudo-covariance of a signal even when utilizing the quadratic error function. This model of adaptive transversal filter is referred to as the widely linear filter.

VI. Complex Augmented Transversal Filter

In Section II-C, we explored the transversal filter with complex coefficients, demonstrating its capability to handle complex signals in various configurations, although an explicit iterative solution was not provided. It is important to note that, despite its versatility, this filter design does not account for the statistical characteristics associated with the non-circularity exhibited by the signals, a point that has been emphasized throughout. To address this limitation, we introduce a novel approach, rather than merely extending the real transversal filter to the complex domain.

This leads us to the concept of the *widely linear filter*, a pivotal design for the analysis of complex signals. It provides an expression that incorporates both covariance and pseudo-covariance matrices, addressing the limitations of the conventional transversal filter. With this filter model and the use of augmented notation, we can derive expressions similar to those obtained in conventional complex filtering (Section IV-F). These expressions yield optimal coefficients and will be demonstrated at the end of this section.

A. The Widely Linear modeling

One way to deal with pseudo-correlations that may be present in certain types of signals is to use a different scheme than what has been employed before. The chosen design is a configuration composed of a pair of transversal filters working together, as shown in Fig. 16. The blocks indicated by $\mathbf{g}$ and $\mathbf{h}$ are complex transversal filters with size M each, as described in Fig. 4 (abstracted in Fig. 5). The block indicated by the dashed line in the figure represents the composite filter, denoted here by the symbol $\mathbf{w}_c$. This configuration is called

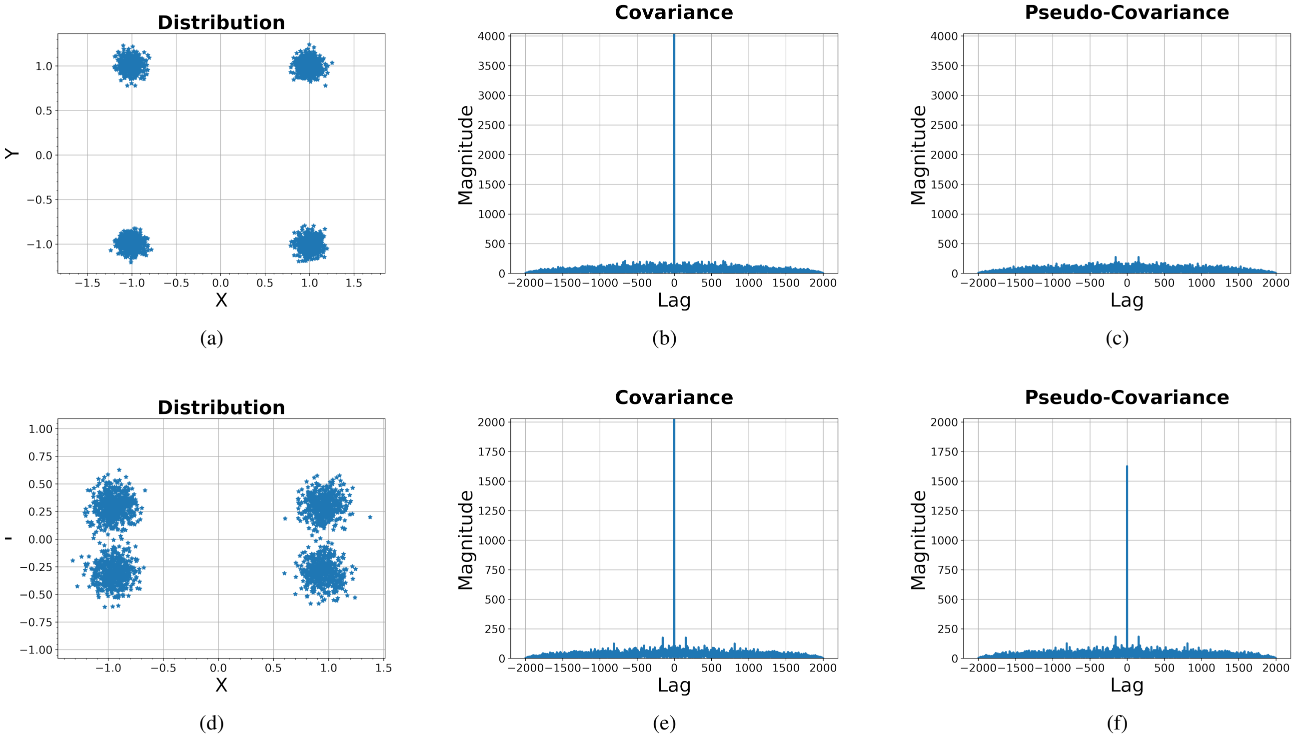


Fig. 15. Distribution of 4QAM in two cases:(a) circular and (d) non-circular. In (b), The covariance of the circular signal version. (c) The pseudo-covariance of the circular version of signal. (e) The covariance of the non-circular signal. (f) The pseudo-covariance of the non-circular version of signal.

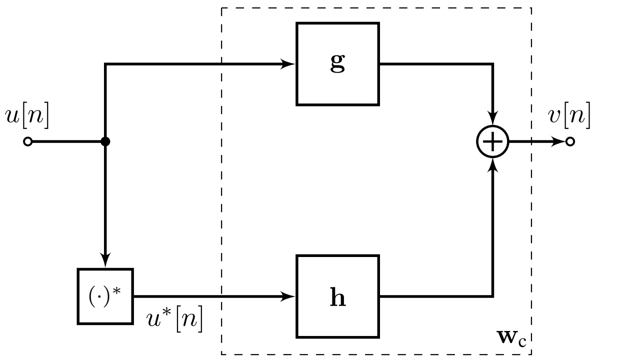


Fig. 16. Configuration of a widely linear transversal filter. The input signal is split into its non-conjugated component, processed by the transversal filter $\mathbf{g}$, and its conjugated component, obtained via the $(\cdot)^*$ block, processed by the transversal filter $\mathbf{h}$. The outputs are summed to produce the signal $v[n]$.

a *widely linear* filter, and is named so because of the form of its output,

$$v[n] = \mathbf{g}^H \mathbf{u}[n] + \mathbf{h}^H \mathbf{u}^*[n]. \quad (45)$$

Equation (45) represents a linear combination of the input signal and its conjugate, utilizing filter coefficients. Note that this expression is linear in $\mathbb{C}^{2M}$ for the variables $\mathbf{w}_c$ and $\underline{\mathbf{u}}[n]$, but non-linear in the signal space $\mathbb{C}^M$, where $\mathbf{u}[n]$ resides. This non-linearity arises from the conjugation operation, as shown in Fig. 16.

Remember that the two complex transversal filters which compose the system can be represented by vectors with complex coefficients, $\mathbf{g} = \begin{bmatrix} g_0 & g_1 & \dots & g_{M-1} \end{bmatrix}^T$ and $\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \dots & h_{M-1} \end{bmatrix}^T$ with $g_k, h_l \in \mathbb{C}$, $\forall k, l \in \{0, 1, \dots, M-1\}$.

Similarly to what was done in Fig. 5, the filter design can be abstracted into a box that provides a widely linear output as shown in Fig. 17. When the structure is represented in this

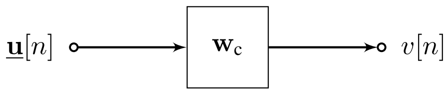


Fig. 17. Abstraction design for a widely linear filter. The augmented signal $\underline{\mathbf{u}}[n]$ is filtered by a pair of transversal filters, represented by the box $\mathbf{w}_c$, resulting into a scalar signal $v[n] = \mathbf{g}^H \mathbf{u}[n] + \mathbf{h}^H \mathbf{u}^*[n]$.

way, it is important to make it clear which configuration is used at each moment to avoid confusion. In order to avoid this in this work, we will always state which structure is used at the moment: the usual one for the transversal structure in Fig. 4 and the widely linear one for the one shown in Fig. 16.

As we conclude this topic, a pertinent question may arise: can the widely linear filter be effectively employed in the configurations illustrated in Fig. 7, Fig. 8, Fig. 9, and Fig. 10, among others? The answer is affirmative, albeit with considerations: the widely linear filter proves particularly useful when pseudo-correlation is a significant concern. In scenarios where pseudo-correlation is less critical, the transversal filter may suffice.

Before examining the application of the widely linear filter in the aforementioned configurations, it is crucial to understand how the optimal coefficients are calculated concerning the given input and output. This aspect will be explored in the subsequent section.

B. The Widely Linear Optimal Solution

The widely linear filter design shown in Fig. 16 produces a signal output $v[n]$ given by expression (45). This output can be used, for example, to minimize the mean square error in the same way as was done in Section IV-F to obtain the optimum filter coefficients. Before that, it is important to compress the widely linear output expression into a shorter notation,

$$v[n] = \begin{bmatrix} \mathbf{g}^H & \mathbf{h}^H \end{bmatrix} \begin{bmatrix} \mathbf{u}[n] \\ \mathbf{u}^*[n] \end{bmatrix} = \mathbf{w}_c^H \underline{\mathbf{u}}[n].$$

Here, $\mathbf{w}_c^H = \begin{bmatrix} \mathbf{g}^H & \mathbf{h}^H \end{bmatrix}$ is a compose complex vector and $\underline{\mathbf{u}}[n]$ is the augmented complex vector of the input signal. The subscript 'c' will be used to name composite vectors that are constructed by placing one complex vector over the other. Note that the augmented vector is a particular case of a composite vector and should not be confused with it.

The expression for the filter error in relation to a training signal $d[n]$ is

$$e_{\text{WL}}[n] = d[n] - \mathbf{w}_c^H \underline{\mathbf{u}}[n].$$

Here, the WL subscript emphasizes that the error is derived from a widely linear filter.

With the error expression in hand, we can define a cost function that represents the mean squared error, analogous to (4). The widely linear mean-square error is given by

$$\mathcal{J}_{\text{WLMSE}} = \text{E}\{(d[n] - \mathbf{w}_c^H \underline{\mathbf{u}})(d[n] - \mathbf{w}_c^H \underline{\mathbf{u}})^*\}. \tag{46}$$

To develop the previous equation, it is convenient to define, in addition to the scalar variance already mentioned $\sigma_d^2 = E\{|d[n]|^2\}$, the expressions for:

- The complex cross-covariance composite vector is

$$\mathbf{p}_c[n] = E\left\{ d^*[n] \begin{bmatrix} \mathbf{u}[n] \\ \mathbf{u}^*[n] \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{p}[n] \\ \mathbf{q}^*[n] \end{bmatrix},$$

  where $\mathbf{p}[n] = E\{d^*[n]\mathbf{u}[n]\}$, $\mathbf{q}[n] = E\{d[n]\mathbf{u}[n]\}$.
- The augmented covariance matrix for the input signal is

$$\underline{\mathbf{C}}[n] = \begin{bmatrix} \mathbf{C}[n] & \tilde{\mathbf{C}}[n] \\ \tilde{\mathbf{C}}^*[n] & \mathbf{C}^*[n] \end{bmatrix},$$

  with $\mathbf{C}[n] = E\{\mathbf{u}[n]\mathbf{u}^H[n]\}$ and $\tilde{\mathbf{C}}[n] = E\{\mathbf{u}[n]\mathbf{u}^T[n]\}$.

Considering the statistics of a WSS process in which the correlations do not vary with time, the obtained cost function (46) can be written as

$$\mathcal{J}_{\text{WLMSE}} = \sigma_d^2 - \mathbf{p}_c^H \mathbf{w}_c - \mathbf{w}_c^H \mathbf{p}_c + \mathbf{w}_c^H \underline{\mathbf{C}} \mathbf{w}_c \ . \tag{47}$$

The expression obtained in (47) has the same format as the complex expression in (28). With this, we can obtain the expressions for the derivative (gradient) and the optimal solution (in terms of the minimum mean square error).

Recall that the set of deterministic coefficients is a composite complex variable that can be considered as a regular complex variable with twice the dimension. By differentiating the previous expression (47) with respect to this variable (Table I), one obtains the expression for the adaptive gradient,

$$\frac{\partial \mathcal{J}_{\text{WLMSE}}}{\partial \mathbf{w}_c^H} = -\mathbf{p}_c + \underline{\mathbf{C}} \mathbf{w}_c \tag{48}$$

The gradient equation is noticed to have the same form as the one developed for the usual complex case (29). With regard to the components, the reader can easily verify that the previous equation is equivalent to the system,

$$\begin{cases} \frac{\partial \mathcal{J}_{\text{WLMSE}}}{\partial \mathbf{g}^H} = -\mathbf{p} + \mathbf{C}\mathbf{g} + \tilde{\mathbf{C}}\mathbf{h} \\ \frac{\partial \mathcal{J}_{\text{WLMSE}}}{\partial \mathbf{h}^H} = -\mathbf{q}^* + \tilde{\mathbf{C}}^*\mathbf{g} + \mathbf{C}^*\mathbf{h} \end{cases} . \tag{49}$$

The expression for the optimal coefficients, which is the zero of (48), will also maintain the format of its non-augmented version and can be written by

$$\mathbf{w}_c^{\text{opt}} = \underline{\mathbf{C}}^{-1} \mathbf{p}_c. \tag{50}$$

The same optimal solution could be found while solving the system equation in (49). The compose vector notation adopted for the augmented filtering not only saves ink but also allows for easy algebraic manipulation, as long as care is taken when using the relationships developed in section IV-A.

Note that the optimal widely linear solution has the same form as the optimal complex linear expression in (30), but now takes into account pseudo-correlation as well as composite cross-correlation. The optimal solution here presents the same problem as the one found in the linear case: the presence of a matrix inversion. In the augmented case, the operation becomes more costly due to the dimension of the augmented correlation matrix $\underline{\mathbf{C}}$, which is twice that of the complex correlation $\mathbf{C}$ (if the size M of the linear filter $\mathbf{w}$ is kept the same for each filter $\mathbf{g}$, $\mathbf{h}$ that composes the widely linear filter). To overcome the inversion obstacle, the next topic discusses some methods used for the complex case that can be adapted for the widely linear case.

VII. Iterative methods for transversal filters

In Sections IV-F and VI-B, optimal solutions for both the complex and augmented versions of the transversal filter were developed. In both cases, it would be highlighted that an exact solution was infeasible due to the computationally costly matrix inversion required. The importance of the gradient, which represents the derivative of the cost function, was also emphasized. To mitigate the computational costs associated with the computation of matrix inversion, iterative techniques are often preferred [68], [69]. This section explores some principles of these techniques to address these challenges.

The iterative approach begins with an initial estimate and proceeds to perform a series of calculations to refine the estimate until specific stopping criteria are met. Given a complex vector $\mathbf{w}$ of arbitrary size ($\mathbf{w}_c$ for the WL filter or $\mathbf{w}$

for a linear complex filter, for example), an iterative process of first order can be expressed using a difference equation,

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \Delta\mathbf{w}[n]. \tag{51}$$

Starting from an arbitrary initial value $\mathbf{w}[0]$, the method successively iterates over n until the stopping condition is satisfied. $\Delta\mathbf{w}[n]$ represents the update that should be applied at each iteration. What differentiates a specified algorithm from the other is how the update is calculated.

This principle can be extended directly to the augmented case with

$$\underline{\mathbf{w}}[n+1] = \underline{\mathbf{w}}[n] + \Delta\underline{\mathbf{w}}[n]. \tag{52}$$

The augmented notation better encapsulates the definitions that we will see for the gradients, making it the preferred choice.

Algorithms that utilize the gradient as an adjustment factor typically express the increment factor as

$$\Delta\underline{\mathbf{w}}[n] = -2\mu[n]\underline{\mathbf{M}}^{-1}(\underline{\mathbf{w}}[n])\frac{\partial f}{\partial \underline{\mathbf{w}}^H}(\underline{\mathbf{w}}[n]). \tag{53}$$

Here, $\underline{\mathbf{M}}(\underline{\mathbf{w}}[n])$ is an augmented matrix, and $\partial f/\partial\underline{\mathbf{w}}^H(\underline{\mathbf{w}}[n])$ denotes the gradient of the cost function f at the point $\underline{\mathbf{w}}[n]$. The real scalar $\mu[n]$ that adjusts the size of the increment for each iteration is called the step (or learning) factor.

The algorithms that employ the matrix $\underline{\mathbf{M}}$ can be classified into two main categories, depending on the choice of this matrix:

- If $\underline{\mathbf{M}}$ is equal to the identity matrix $\mathbf{I}$, the algorithm is known as the gradient descent method.
- When $\underline{\mathbf{M}}$ is equal to the augmented Hessian matrix $\underline{\mathbf{H}}$ of the cost function, the algorithm is known as the Newton method. The Hessian represents the second derivative and is used to obtain curvature information of the surface curvature of the cost function.

Fig. 18 illustrates this categorization as a set map. Recall that these methods can contain a multiplicative constant $\mu = \mu[n]$, assumed constant from now on without loss of generality. A too large step size can lead to oscillations around the optimal solution, while a too small step size can result in slow convergence of the algorithm.

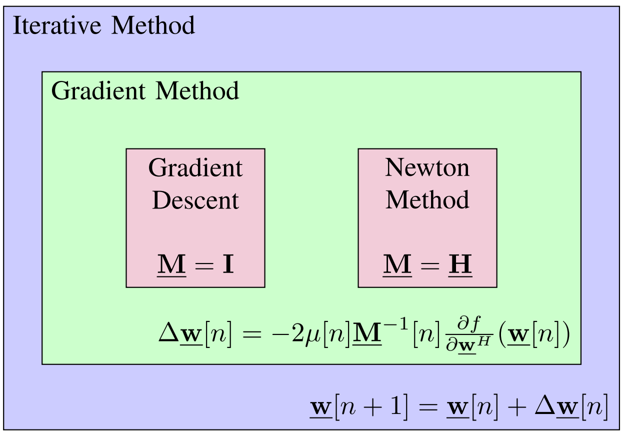


Fig. 18. Sets that categorize the main methods utilizing the gradient as a parameter adjustment method.

A. The General Gradient Augmented Expression

Here we develop the iterative expression (52) and the adjustment factor (53) based on equivalent real methods using the relationships described in Section IV-A.

Let $\mathbf{r} = \begin{bmatrix}\mathbf{x}^T & \mathbf{y}^T\end{bmatrix}^T$ be a real $\mathbb{R}^{2M}$ compose vector, the general method for optimizing a real function f is given by the iterative method.

$$\mathbf{r}[n+1] = \mathbf{r}[n] - \mu\mathbf{M}^{-1}(\mathbf{r}[n])\frac{\partial f}{\partial \mathbf{r}^T}(\mathbf{r}[n]). \tag{54}$$

in which $\mathbf{M}(\mathbf{r}[n])$ (without underline) is a real positive definite matrix at point $\mathbf{r}[n]$. By multiplying both sides in (54) from left by the matrix $\mathbf{T}$ (11) we obtain the expression

$$\mathbf{Tr}[n+1] = \mathbf{Tr}[n] - \mu\mathbf{TM}^{-1}(\mathbf{r}[n])\frac{\partial f}{\partial \mathbf{r}^T}(\mathbf{r}[n]). \tag{55}$$

From now on, we will avoid overloading the notation used, and for this purpose, any term without the specified point will be considered as in $\mathbf{w}[n]$ for augmented and in $\mathbf{r}[n]$ in the real compose case. Applying the gradient transformation (20) to the expression (55) we can write the update rule

$$\mathbf{Tr}[n+1] = \mathbf{Tr}[n] - \mu\mathbf{TM}^{-1}\mathbf{T}^H\frac{\partial f}{\partial \underline{\mathbf{w}}^H}. \tag{56}$$

Finally, applying the coordinate transformation relation (10) and also through the matrix relation between real and augmented matrices $\underline{\mathbf{M}} = \frac{1}{2}\mathbf{TM}^{-1}\mathbf{T}^H$ [28], the augmented version of the update rule (54) can be expressed as

$$\underline{\mathbf{w}}[n+1] = \underline{\mathbf{w}}[n] - 2\mu\underline{\mathbf{M}}^{-1}\frac{\partial f}{\partial \underline{\mathbf{w}}^H}. \tag{57}$$

It is important to note that Equation (57) represents an augmented expression: There is implicit redundancy, since one component is the conjugate of the other. To make it clearer, let the inverse of the matrix $\underline{\mathbf{M}}$, at some arbitrary valid point, be given by [28]

$$\underline{\mathbf{M}}^{-1} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^* & \mathbf{A}^*\end{bmatrix}. \tag{58}$$

Using (58) above in Equation (57) and taking the first component, we have

$$\mathbf{w}[n+1] = \mathbf{w}[n] - 2\mu\left(\mathbf{A}\frac{\partial f}{\partial \mathbf{w}^H} + \mathbf{B}\frac{\partial f}{\partial \mathbf{w}^T}\right), \tag{59}$$

which is the expression of the complex (non-augmented) gradient. Note that the expressions (57) and (59) are equivalent. It is important to note that while the augmented version is redundant, it also becomes more compact, which is why it was used in Fig. 18.

The multiplicative factor of two that appears in the expressions because of the form of the Wirtinger derivatives can be incorporated into the step factor or not. Care must be taken when choosing how to compare it with complex algorithms that may or may not incorporate this factor [70], [71].

B. The Gradient Descent

The gradient descent, as shown in Fig. 18, is the case where the matrix $\underline{\mathbf{M}}$ used is the identity. Using directly the equation (59) one can determine the expression for the stochastic gradient for two situations: the complex case and the widely linear case. The function for which the gradients are to be calculated from now on is the cost function.

- For the case of the complex linear filter $\mathbf{w}[n] \in \mathbb{C}^M$, where M is the size of transversal filter, the gradient descent update is given by

$$\mathbf{w}[n+1] = \mathbf{w}[n] - 2\mu \frac{\partial \mathcal{J}}{\partial \mathbf{w}^H}. \tag{60}$$

- In the widely linear case $\mathbf{w}_\mathrm{c}$, the update rule becomes

$$\mathbf{w}_\mathrm{c}[n+1] = \mathbf{w}_\mathrm{c}[n] - 2\mu \frac{\partial \mathcal{J}}{\partial \mathbf{w}_\mathrm{c}^H}. \tag{61}$$

Note that in the case of the widely linear filter, $\mathbf{w}_\mathrm{c}$ is still a regular complex variable (non-augmented), and in both cases, the identity matrix is used, maintaining the same format for the solutions.

Replacing equation (48) into (61), we have the expression for the cost function $\mathcal{J}_{\mathrm{WLMSE}}$ as

$$\mathbf{w}_\mathrm{c}[n+1] = \mathbf{w}_\mathrm{c}[n] + 2\mu(\mathbf{p}_\mathrm{c}[n] - \underline{\mathbf{C}}[n]\mathbf{w}_\mathrm{c}[n]). \tag{62}$$

The linear CMSE solution has the same form as the expression above, simply replacing $\mathbf{w}_\mathrm{c}$ with $\mathbf{w}$, $\mathbf{p}[n] = \mathbf{p}_\mathrm{c}[n]$ and $\underline{\mathbf{C}}[n]$ with $\mathbf{C}[n]$.

The expression (62) in terms of its components

$$\mathbf{g}[n+1] = \mathbf{g}[n] + 2\mu(\mathbf{p} - \mathbf{C}\mathbf{g}[n] - \tilde{\mathbf{C}}\mathbf{h}[n])$$
$$\mathbf{h}[n+1] = \mathbf{h}[n] + 2\mu(\mathbf{q}^* - \tilde{\mathbf{C}}^*\mathbf{g}[n] - \mathbf{C}^*\mathbf{h}[n]).$$

The entire development could be carried out for $\mathbf{g}$ and $\mathbf{h}$ separately, but writing in terms of extended and augmented vectors makes the development less tedious while maintaining the solution similar to the usual case.

C. Values for the Step Size

The step size, denoted by μ, is a crucial parameter in gradient descent and various other optimization algorithms. It regulates the size of parameter updates in each iteration. There are two main approaches to the step size. In the first approach, the step size remains constant throughout the optimization process. In the second approach, the step size values can be adjusted during the coefficient adjustment process. In this topic, we will discuss the different values that can be assigned to the step size in the context of the widely linear filter. The appropriate choice of this parameter is critical, as it directly affects the speed and effectiveness of algorithm convergence.

To analyse the step size, we will rewrite the gradient descent equation (62) as follows:

$$\mathbf{w}_\mathrm{c}[n+1] = (\mathbf{I} - \mu\underline{\mathbf{C}})\,\mathbf{w}_\mathrm{c}[n] + \mu\mathbf{p}_\mathrm{c}.$$

Here, the identity matrix $\mathbf{I}$ has a size of $2M \times 2M$. By shifting the filter coefficients in relation to the optimal solution in the form of $\mathbf{v}_\mathrm{c}[n+1] = \mathbf{w}_\mathrm{c}[n+1] - \mathbf{w}_\mathrm{c}^{\mathrm{opt}}$ the previous equation can be written as

$$\mathbf{v}_\mathrm{c}[n+1] = (\mathbf{I} - \mu\underline{\mathbf{C}})\,\mathbf{v}_\mathrm{c}[n].$$

Diagonalizing the covariance matrix as $\underline{\mathbf{C}} = \mathbf{Q}\underline{\Lambda}\mathbf{Q}^H$ and multiplying both sides by the eigenvector matrix $\mathbf{Q}^H$ we obtain

$$\mathbf{Q}^H\mathbf{v}_\mathrm{c}[n+1] = \left(\mathbf{Q}^H\mathbf{Q} - \mu\mathbf{Q}^H\mathbf{Q}\underline{\Lambda}\right)\mathbf{Q}^H\mathbf{v}_\mathrm{c}[n]$$

Lastly, we rename $\mathbf{Q}^H\mathbf{v}_\mathrm{c}[n] = \boldsymbol{\eta}_\mathrm{c}[n]$, allowing for a simplified form of the expression,

$$\boldsymbol{\eta}_\mathrm{c}[n+1] = (\mathbf{I} - \mu\underline{\Lambda})\,\boldsymbol{\eta}_\mathrm{c}[n]. \tag{63}$$

In order to ensure that the iterative method remains convergent, the idea is that during each iteration, the magnitude of the updates in the parameters $\boldsymbol{\eta}_\mathrm{c}$ does not increase. To achieve this, the adopted condition is that $|1 - \mu\underline{\lambda}_k| < 1$, for $\underline{\Lambda} = \mathrm{diag} = \{\underline{\lambda}_k\}_{k=1}^{2(M+1)}$. With the condition that the step size should be positive to avoid changing the direction of the gradient, we can write the stability condition as

$$0 < \mu < \frac{2}{\underline{\lambda}_{\mathrm{max}}}$$

where $\underline{\lambda}_{\mathrm{max}}$ is the maximum eigenvalue of the augmented covariance matrix.

Since computing the eigenvalues can be computationally expensive, the trace of the matrix is often used. With this, the stability condition for the step size can be expressed as

$$0 < \mu < \frac{2}{\underline{\lambda}_{\mathrm{max}}} < \frac{2}{\mathrm{tr}(\underline{\mathbf{C}})}, \tag{64}$$

due the relation

$$\mathrm{tr}(\underline{\mathbf{C}}) = \sum_{k=1}^{2(M+1)} \underline{\lambda}_k \geq \underline{\lambda}_{\mathrm{max}}.$$

The operator tr denotes the trace of the matrix. This condition is more restrictive while also ensuring the algorithm's stability concerning the step size factor.

The same steps can be applied to the non-augmented complex case, resulting in

$$0 < \mu < \frac{2}{\mathrm{tr}(\mathbf{C})} \quad \text{(complex case)}.$$

As an exercise, the reader can easily verify this equation.

It is important to note that the maximum step size required by the widely linear approximation should be smaller than its linear counterpart if each length of the widely linear case is the same as that of the linear case. This is because the trace of the augmented matrix $\underline{\mathbf{C}}$ is greater than the trace of $\mathbf{C}$, resulting in a decreased convergence speed to achieve stability in the augmented case.

D. The Newton Method

The Newton method is an optimization algorithm that uses second-order information provided by the cost function. The method takes into account the curvature of the function, encoded as the Hessian matrix, helping to find the optimal point.

The advantage of Newton's method is its ability to approach a critical point more efficiently than methods such as stochastic gradient descent. The downside is the computational cost involved in the calculation and inversion of the Hessian matrix. For widely linear filters, Newton's method is based on the augmented correlation matrix and can be obtained by the second-order approximation of the Taylor series of the cost function.

We will employ the second-order approximation of the Taylor series to demonstrate that the Hessian matrix is equivalent to $\underline{\mathbf{M}}$. Additionally, we will express Hessian using the second-order derivative notation as described in (23).

1) Optimal increment with Taylor's series: The Taylor series for a real function f in its second-order approximation in the composite real case $\mathbf{r} = \begin{bmatrix}\mathbf{x}^T & \mathbf{y}^T\end{bmatrix}^T$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$ around a point $\mathbf{r}$ is given by

$$f(\mathbf{r}+\mathbf{k}) \approx q = f + \mathbf{k}^T \frac{\partial f}{\partial \mathbf{r}^T} + \frac{1}{2}\mathbf{k}^T \mathbf{H}\mathbf{k}. \tag{65}$$

In the previous expression, q refers to this being a quadratic approximation, $\mathbf{k} \in \mathbb{R}^{2M}$ is the increment vector to be used and the matrix $\mathbf{H}$ is the Hessian matrix relative to the function f in terms of the compose vector $\mathbf{r}$, given by (21). With expression (65) it is possible to obtain the augmented version using the relationships described in IV-A, which is

$$q = f + \underline{\boldsymbol{\delta}}^H \frac{\partial f}{\partial \underline{\mathbf{z}}^H} + \frac{1}{2}\underline{\boldsymbol{\delta}}^H \underline{\mathbf{H}}\underline{\boldsymbol{\delta}}, \tag{66}$$

with

$$\underline{\mathbf{H}} = \frac{1}{4}\mathbf{T}^H \mathbf{H}\mathbf{T}, \tag{67}$$

and $\underline{\boldsymbol{\delta}} = \mathbf{T}\mathbf{k}$.

To determine the form of $\underline{\mathbf{H}}$, we simply need to expand the expression in (67). With this,

$$\begin{aligned}\underline{\mathbf{H}} &= \frac{1}{4}\begin{bmatrix}\mathbf{I} & j\mathbf{I}\\ \mathbf{I} & -j\mathbf{I}\end{bmatrix}\begin{bmatrix}\frac{\partial}{\partial \mathbf{x}^T}\cdot\frac{\partial f}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{x}^T}\cdot\frac{\partial f}{\partial \mathbf{y}}\\ \frac{\partial}{\partial \mathbf{y}^T}\cdot\frac{\partial f}{\partial \mathbf{x}} & \frac{\partial}{\partial \mathbf{y}^T}\cdot\frac{\partial f}{\partial \mathbf{y}}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{I}\\ -j\mathbf{I} & j\mathbf{I}\end{bmatrix}\\ &= \begin{bmatrix}\frac{\partial}{\partial \mathbf{w}^H}\cdot\frac{\partial f}{\partial \mathbf{w}} & \frac{\partial}{\partial \mathbf{w}^H}\cdot\frac{\partial f}{\partial \mathbf{w}^*}\\ \frac{\partial}{\partial \mathbf{w}^T}\cdot\frac{\partial f}{\partial \mathbf{w}} & \frac{\partial}{\partial \mathbf{w}^T}\cdot\frac{\partial f}{\partial \mathbf{w}^*}\end{bmatrix} = \frac{\partial}{\partial \underline{\mathbf{w}}^H}\cdot\frac{\partial f}{\partial \underline{\mathbf{w}}},\end{aligned} \tag{68}$$

using the adopted notation. The factor of 1/4 is absorbed by the Wirtinger operators to form the augmented Hessian. With the Hessian matrix determined, we can continue the development through the Taylor series. The optimal value for $\underline{\boldsymbol{\delta}}$ is obtained by the zero of the derivative with respect to the increment of the expression (66),

$$\frac{\partial q}{\partial \underline{\boldsymbol{\delta}}^H} = \frac{\partial f}{\partial \underline{\mathbf{w}}^H} + \frac{1}{2}\underline{\mathbf{H}}\underline{\boldsymbol{\delta}} = 0.$$

The optimal increment is

$$\underline{\boldsymbol{\delta}} = -2\underline{\mathbf{H}}^{-1}\frac{\partial f}{\partial \underline{\mathbf{w}}^H},$$

which is the same increment $\underline{\boldsymbol{\delta}} = \Delta\underline{\mathbf{w}}$ described in Fig. 18, in the case of the Newton method.

2) The method applied to linear filter: The previous definitions can be applied to the complex linear case using the cost function given by (28). The components of the Hessian augmented matrix $\underline{\mathbf{H}}$, in relation to the coefficients $\mathbf{w}$, for this cost function are given by

$$\frac{\partial}{\partial \mathbf{w}^H}\cdot\frac{\partial}{\partial \mathbf{w}}\mathcal{J}_{\text{CMSE}} = \mathbf{C}$$

and

$$\frac{\partial}{\partial \mathbf{w}^H}\cdot\frac{\partial}{\partial \mathbf{w}^*}\mathcal{J}_{\text{CMSE}} = 0.$$

The Hessian matrix relative to the function $\mathcal{J}_{\text{CMSE}}(\mathbf{w})$ is

$$\underline{\mathbf{H}} = \begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^*\end{bmatrix}, \tag{69}$$

allowing us to write the iterative expression for the filter case, with the presence of a step factor, as

$$\mathbf{w}[n+1] = \mathbf{w}[n] - 2\mu\mathbf{C}^{-1}\frac{\partial}{\partial \mathbf{w}^H}\mathcal{J}_{\text{CMSE}}. \tag{70}$$

3) The method applied to widely linear filter: The same procedure can be done for the widely linear filter case considering the cost function as WLMSE. The components of the Hessian augmented matrix $\underline{\mathbf{H}}_{\text{c}}$, in relation to the coefficients $\mathbf{w}_c$, for the cost function (47) are given by the expressions

$$\frac{\partial}{\partial \mathbf{w}_{\text{c}}^H}\cdot\frac{\partial}{\partial \mathbf{w}_{\text{c}}}\mathcal{J}_{\text{WLMSE}} = \underline{\mathbf{C}}$$

and

$$\frac{\partial}{\partial \mathbf{w}_{\text{c}}^H}\cdot\frac{\partial}{\partial \mathbf{w}_{\text{c}}^*}\mathcal{J}_{\text{WLMSE}} = 0.$$

The Hessian matrix relative to $\mathcal{J}_{\text{WLMSE}}(\mathbf{w}_{\text{c}})$ is

$$\underline{\mathbf{H}}_{\text{c}} = \begin{bmatrix}\underline{\mathbf{C}} & \mathbf{0}\\ \mathbf{0} & \underline{\mathbf{C}}^*\end{bmatrix}. \tag{71}$$

The iterative expression for the widely linear filter case can be expressed as

$$\mathbf{w}_{\text{c}}[n+1] = \mathbf{w}_{\text{c}}[n] - 2\mu\underline{\mathbf{C}}^{-1}\cdot\frac{\partial}{\partial \mathbf{w}_{\text{c}}^H}\mathcal{J}_{\text{WLMSE}}. \tag{72}$$

Here, the gradient is also given by (48).

Linear and augmented complex development are similar because both use the standard vector of a complex variable. However, in the compose case, the complex variable needs to be divided into two other variables, $\mathbf{g}$ and $\mathbf{h}$, implying that $\mathbf{w}_c$ must have an even complex dimension (this requirement is already met by the design of the widely linear filter).

VIII. Stochastic Approximations and Algorithms

The iterative methods presented in the previous section provide an alternative to the optimal solutions proposed in Equations (30) and (50). These optimal solutions suffer from the drawback of matrix inversion, as mentioned earlier. The astute reader may have noticed that Newton's method also involves matrix inversion in its equations, presenting the same issue. Furthermore, all of these methods require computation of the correlation expectations, which is impractical in real-world scenarios. This section aims to address these challenges by transforming the aforementioned iterative methods into algorithms that approximate the values of covariances and matrix inverses.

The first algorithm is the LMS (Least Mean Square) algorithm, which is a version of the gradient method considering real-time data. The last algorithm is the Newton algorithm that utilizes the matrix inversion lemma for inverse estimation, following the same process of estimating correlations as the LMS method.

A. Common Structure of the Algorithms

All the algorithms presented in this section — CLMS, WLLMS, and the augmented Newton method — share the same fundamental structure. They operate on discrete-time signals by using a sliding window of size M. At each iteration, the algorithm uses the current segment of the signal to compute the error and update the filter coefficients. The general steps are as follows:

- **Initialization**
  - Set N as the total number of samples in the input signal.
  - Initialize a vector of coefficients ($\mathbf{w}$ or $\mathbf{w}_c$) with zeros.
  - Initialize an error vector to store the error values at each iteration.
  - (For Newton's method) Initialize a covariance inverse matrix, typically as an identity matrix.
- **Sliding Window Loop**
  - Iterate over the signal from $n = M-1$ to $N-1$, to avoid taking values outside the signal.
  - Build the input vector $\mathbf{u}$, and if needed, the augmented vector $\underline{\mathbf{u}}$.
  - Compute the error between the desired signal and the estimated output.
  - Update the filter coefficients based on the algorithm's rule (gradient step or Newton update).
- **Output**
  - Return the final filter coefficients and the error signal.

This behavior is illustrated in Fig. 19, where each iteration processes a segment of the input signal (represented in blue) and shifts the window one sample to the right. Understanding this common iterative framework is essential, as it serves as the basis for all adaptive filtering algorithms discussed in this work. Readers are encouraged to implement these routines using programming languages such as Julia [72] or Python [73]. In the following sections, each algorithm is presented in details.

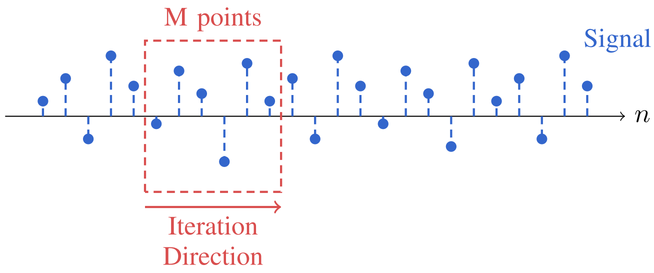


Fig. 19. Example of windowing used in recursive algorithms (CLMS, WLLMS, and Newton). Each iteration takes into account the signal (blue) contained within the red dashed region for computation and after this, the window shifts one sample to the right only stopping at the end of signal.

B. The LMS Algorithm

This topic explores the widely used least mean square (LMS) algorithm, a well-known iterative method that serves as an alternative to obtain optimal solutions without the need for matrix inversions. Here, we will explore both the complex linear (CLMS) and the widely linear (WLLMS) cases of the LMS algorithm.

1) The CLMS Case: This algorithm is directly applicable to the linear configuration illustrated in Fig. (4). The fundamental concept behind LMS lies in approximating covariances through instantaneous approximations in the expression (60). The estimation process is achieved by considering

$$\hat{\mathbf{p}}[n] = d^*[n]\mathbf{u}[n]$$
$$\hat{\mathbf{C}}[n] = \mathbf{u}[n]\mathbf{u}^H[n].$$

The circumflex accent (^) above each correlation indicates that it is an instantaneous estimative. It is important to emphasize that different types of estimation can be applied. However, in the context discussed, instantaneous estimations are chosen because they depend solely on the current iteration instant. With this, we can develop the equation (60) for the cost function $\mathcal{J}_{\text{CMSE}}$ using the approximations as follows:

$$\begin{aligned}\mathbf{w}[n+1] &= \mathbf{w}[n] - 2\mu\left(\hat{\mathbf{p}}[n] - \hat{\mathbf{C}}[n]\mathbf{w}\right) \\ &= \mathbf{w}[n] - 2\mu\mathbf{u}[n]\left(d^*[n] - \mathbf{u}^H\mathbf{w}\right)\end{aligned},$$

with the error $e[n] = d[n] - \mathbf{w}^H\mathbf{u}[n]$. The expression can be simplified to

$$\mathbf{w}[n+1] = \mathbf{w}[n] - 2\mu\mathbf{u}[n]e^*[n], \tag{73}$$

which is known as the CLMS algorithm. The following Algorithm 1 outlines the operation of Complex LMS as a computer function.

2) The WLLMS Case: Due to the similarity between the CLMS and WLLMS algorithms, the derivation in the widely linear case follows the same steps as in the standard linear case, which is greatly simplified by the use of augmented complex notation. Firstly, estimating covariances,

$$\hat{\underline{\mathbf{C}}}[n] = \underline{\mathbf{u}}[n]\underline{\mathbf{u}}^H[n],$$
$$\hat{\mathbf{p}}_c[n] = d^*[n]\underline{\mathbf{u}}[n].$$

These estimates are then used to derive the iterative update as

$$\mathbf{w}_c[n+1] = \mathbf{w}_c[n] - 2\mu\underline{\mathbf{u}}[n]e^*_{\text{WL}}[n]. \tag{74}$$

The resulting WLLMS algorithm extends the CLMS case with a similar structure, as shown in Algorithm 2.

Algorithm 1 The CLMS algorithm.

Input: u, d, μ, M

$N \leftarrow$ total length of signal d (or u)

Initialize the vector $\mathbf{w}$ of zeros with length M
Initialize the vector $\mathbf{e}$ with size $N - M + 1$

for $n = M - 1$ **to** $N - 1$ **do**

$\mathbf{u} = \begin{bmatrix} u[n] & \cdots & u[n-M+1] \end{bmatrix}^T$
$e[n-M+1] = d[n] - \mathbf{w}^H \mathbf{u}$
$\mathbf{w} \leftarrow \mathbf{w} + \mu \mathbf{u} e^*[n-M+1]$
$n \leftarrow n + 1$

end for

return $\mathbf{w}$, $\mathbf{e}$

Algorithm 2 The WLLMS algorithm.

Input: u, d, μ, M

$N \leftarrow$ total length of signal d (or u)

Initialize the vector $\mathbf{w}_c$ of zeros with length $2M$
Initialize the vector $\mathbf{e}_{\text{WL}}$ with size $N - M + 1$

for $n = M - 1$ **to** $N - 1$ **do**

$\mathbf{u} = \begin{bmatrix} u[n] & \cdots & u[n-M+1] \end{bmatrix}^T$
$\underline{\mathbf{u}} = \begin{bmatrix} \mathbf{u}^T & \mathbf{u}^H \end{bmatrix}^T$
$e_{\text{WL}}[n-M+1] = d[n] - \mathbf{w}_c^H \underline{\mathbf{u}}$
$\mathbf{w}_c \leftarrow \mathbf{w}_c + \mu \underline{\mathbf{u}} e_{\text{WL}}^*[N-M+1]$
$n \leftarrow n + 1$

end for

return $\mathbf{w}_c, \mathbf{e}_{\text{WL}}$

The WLLMS operates similarly to the linear case, with the main difference being the inclusion of dual windowing due to the structure of the widely linear filter. A sliding window of size M is applied to both halves of the filter coefficients: the signal segment used in the first half of $\mathbf{w}_c$ is conjugately replicated for the second half.

Following the same procedure based on instantaneous statistics, one can derive the widely linear Newton method. Here, we focus exclusively on the widely linear version.

C. The Widely Linear Newton Method

In Newton's method for inverse estimation, it is feasible to assign an estimated value for the inverse of the augmented correlation matrix at each discrete time instant. To achieve this, the covariance matrix is recursively estimated as

$$\underline{\hat{\mathbf{C}}}[n+1] = \lambda \underline{\hat{\mathbf{C}}}[n] + \underline{\mathbf{u}}[n+1]\underline{\mathbf{u}}^H[n+1], \tag{75}$$

where $0 < \lambda \leq 1$ is the forgetting factor. Using the matrix inversion lemma [74], the inverse of (75) can be expressed as

$$\underline{\hat{\mathbf{C}}}^{-1}[n+1] = \frac{1}{\lambda}\left[\underline{\hat{\mathbf{C}}}^{-1}[n] - \frac{\underline{\hat{\mathbf{C}}}^{-1}[n]\underline{\mathbf{u}}[n+1]\underline{\mathbf{u}}^H[n+1]\underline{\hat{\mathbf{C}}}^{-1}[n]}{\lambda + \underline{\mathbf{u}}^H[n+1]\underline{\hat{\mathbf{C}}}^{-1}[n]\underline{\mathbf{u}}[n+1]}\right], \tag{76}$$

for an arbitrary and non-zero initial value $\underline{\hat{\mathbf{C}}}^{-1}[0]$. This approach allows the inverse to be iteratively estimated, as detailed in Algorithm 3.

Algorithm 3 Augmented Newton's Method

Input: u, d, μ, M

$N \leftarrow$ total length of signal d (or u)

Initialize the vector $\mathbf{w}_c$ of zeros with length $2M$
Initialize the vector $\mathbf{e}_{\text{WL}}$ with size $N - M + 1$
Initialize the matrix $\underline{\hat{\mathbf{C}}}^{-1}$ (as $\mathbf{I}$, for example)

for $n = M - 1$ **to** $N - 1$ **do**

$\mathbf{u} = \begin{bmatrix} u[n] & \cdots & u[n-M+1] \end{bmatrix}^T$
$\underline{\mathbf{u}} = \begin{bmatrix} \mathbf{u}^T & \mathbf{u}^H \end{bmatrix}^T$
$\underline{\hat{\mathbf{C}}}^{-1} \leftarrow \frac{1}{\lambda}\left[\underline{\hat{\mathbf{C}}}^{-1} - \frac{\underline{\hat{\mathbf{C}}}^{-1}\underline{\mathbf{u}}\underline{\mathbf{u}}^H\underline{\hat{\mathbf{C}}}^{-1}}{\lambda + \underline{\mathbf{u}}^H\underline{\hat{\mathbf{C}}}^{-1}\underline{\mathbf{u}}}\right]$
$e_{\text{WL}}[n-M+1] = d[n] - \mathbf{w}_c^H \underline{\mathbf{u}}$
$\mathbf{w}_c \leftarrow \mathbf{w}_c + \underline{\hat{\mathbf{C}}}^{-1} \underline{\mathbf{u}} e_{\text{WL}}^*[n-M+1]$
$n \leftarrow n + 1$

end for

return $\mathbf{w}_c, \mathbf{e}_{\text{WL}}$

The algorithm for the complex linear case is also similar, with the only distinction being the substitution of augmented covariances and composite coefficients with their conventional complex counterparts. This highlights the benefit of employing composite and augmented notation, as it provides a unified framework that covers both linear and widely linear cases.

IX. Simulations and Results

This section displays the results of some simulated tests using the algorithms developed in the previous section. Firstly, it compares the CLMS and WLLMS algorithms, which are correlated. Then, it compares the complex and widely linear Newton methods. These examples were chosen because they highlight the advantages and disadvantages of each algorithm. The code used in this section was written in Julia and the step size considered already incorporates a factor of 2.

A. The CLMS and WLLMS simulations

The first set of simulations is inspired by the example used in [49], a system identification with some minor alterations. The signal used is modeled by the expression (42). The second example is a noise canceling with a Gaussian and periodic signal.

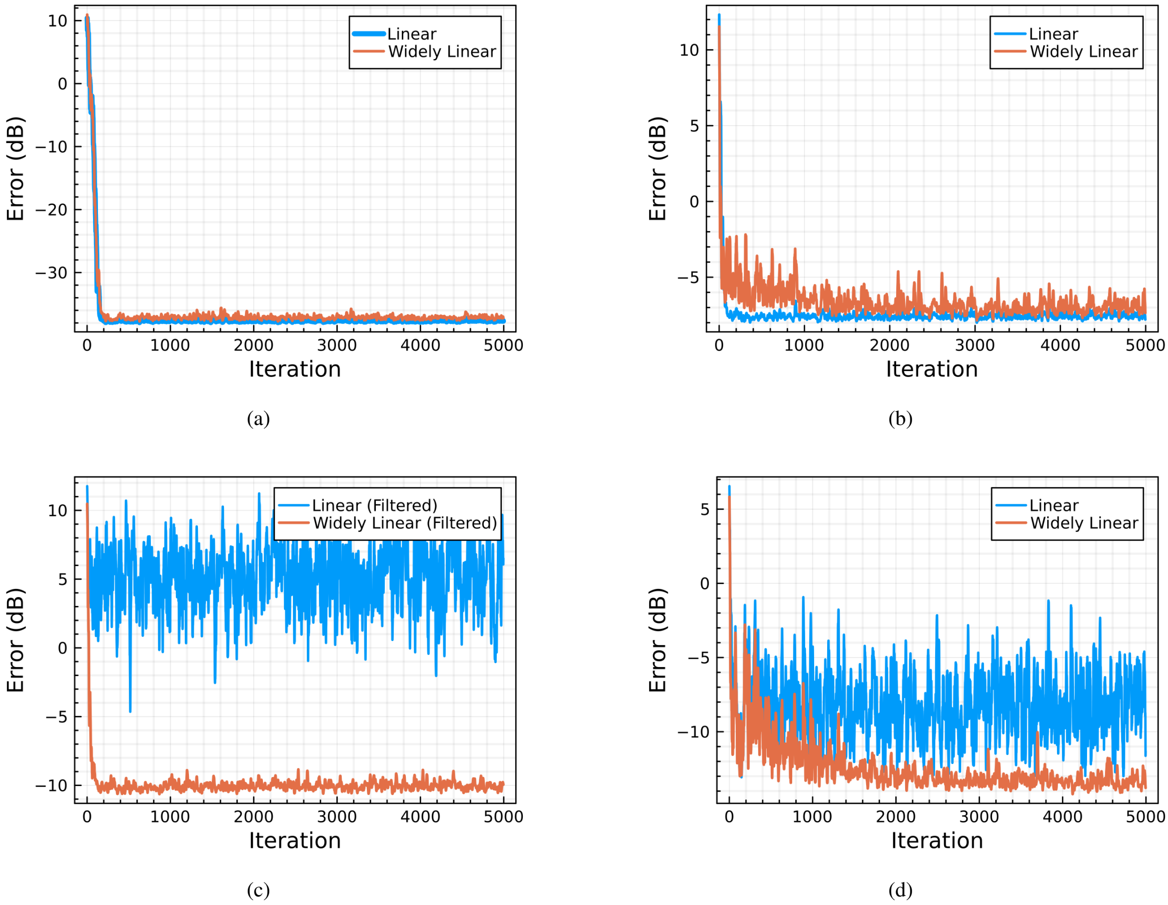


Fig. 20. Set of simulations performed for system identification, where both the signal and noise followed a Gaussian distribution with a signal-to-noise ratio of 20 dB. The analysis was conducted using the CLMS and WLLMS algorithms with a learning rate of $\mu = 0.04$, applied to a dataset of 5×10^3 samples. (a) shows the circular case, while (b) presents the non-circular case. (c) and (d) illustrate the simulations for $d[n] = \text{Re}\{w_{\text{channel}} \star u[n]\}$, corresponding to the circular and non-circular cases, respectively.

1) System identification: The first example is a system identification, as described in Section III-A. The desired signal is generated through a Gaussian process in which the components $A[n]$ and $B[n]$ form the signal according to (42). The signal-to-noise ratio is 20 dB and the step size used was $\mu = 0.04$. The channel to be identified is given by $w_{\text{channel}} = \alpha(1+\cos(2\pi(n-3)/5))-j(1+\cos(2\pi(n-3)/10))$ for $n = 1, \ldots, 5$. The designated filter length is $M = 5$ (in the widely linear case $M = 5$ for each $\mathbf{g}$ and $\mathbf{h}$, resulting in a filter of total length of 10) to match the size of the channel. The complex signals used consisted of the following two cases:

- **Circular case:** The signal has a profile ratio of $\rho = 1/\sqrt{2}$, a circular case, as described in Section V-C. The result of applying the algorithms is shown in Fig. 20(a).
- **Non-Circular case:** The second example illustrated by Fig. 20(b) depicts the scenario where $\rho = 0.1$ and $\varphi = \pi/4$.
  The resulting coefficients at the end of the simulation for the circular case are shown in Fig. 21 for both algorithms. It is important to note the peculiarities in identification using the widely linear filter, which identifies the coefficients but, as it has twice the size of the linear filter, presents an additional solution component (around zeros, as the initial value chosen for the filter coefficients in the algorithm was zero).

In the simulations conducted, there has been no improvement in using the widely linear filter in terms of mean squared error, even for a severe non-circular profile ($\rho = 0.1$). In the latter scenario, the widely linear filter exhibited even poorer performance. Performance improvement occurs in the scenario where the output signal of the channel to be identified is given by $d[n] = \text{real}\{w_{\text{channel}} \star u[n]\}$, where $\star$ represents the complex convolution operation. Keeping the same parameters as before, we obtained the results shown in Fig. 20(c) for the circular case and Fig. 20(d) for the non-circular case. In this case, the advantage of using a widely linear filter is notable.

This example demonstrates that the use of a widely linear filter is not always advantageous, even for a non-circular profile. Each case should be studied and designed, avoiding solely relying on pseudo-covariance. The widely linear filter may perform worse than the linear filter in some cases and better in others, as shown here.

2) Adaptive Signal Estimation: This scenario involves recovering a desired complex signal from a noisy observation and is closely related to system identification, where the system is modeled as a trivial channel corrupted by noise. The focus here is on denoising rather than modeling complex channel dynamics. This simplification is particularly advantageous for directly evaluating the performance of the CLMS and WLLMS algorithms, as it isolates the effects of noise and non-circularity.

In this example, the WLLMS algorithm demonstrates better performance when the desired signal $d[n]$ is derived from the non-linear operation of taking the real part, as shown in Fig. 22. Specifically, a non-circular Gaussian noise is added to a signal of the form $d[n] = \text{Re}\{w \star u[n]\}$, and the adaptive filters aim to estimate $d[n]$ from the noisy observation.

The parameters used were kept consistent with previous experiments, except for the circularity factor, which was set to $\rho = 0.4$ in this case to emphasize the non-circular nature of the signal.

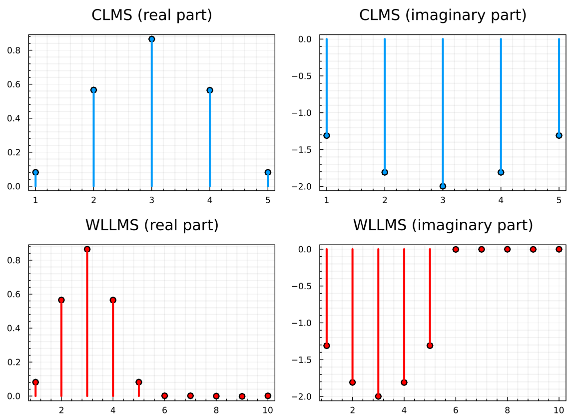


Fig. 21. Analysis of complex circular Gaussian signals with circular noise: A graphical representation displaying the filter coefficients at the conclusion of each algorithm simulation for CLMS and WLLMS.

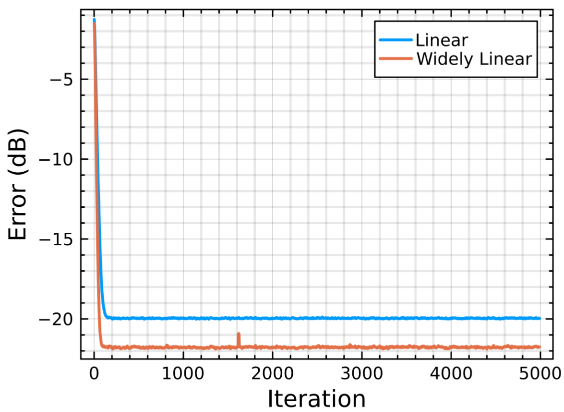


Fig. 22. Adaptive signal estimation for a complex non-circular Gaussian signal with non-circular noise. Plot of the error (in dB) for complex linear (CLMS) and widely linear cases (WLLMS).

The next section presents a similar analysis using the Newton algorithm.

B. *The Newton Method simulations*

The Newton method used in this analysis follows the augmented complex formulation presented in Algorithm 3. Its performance (shown in red) is compared with the WLLMS algorithm (light blue) and the standard complex-valued Newton algorithm (green).

The simulations in this section used the same parameters as those employed in the system identification (Section IX-A2).

Fig. 23 illustrates adaptive signal estimation using Newton's method with non-circular complex signals and noise. The performance of the widely linear filter is slightly inferior (in terms of error minimization) compared to the complex linear case.

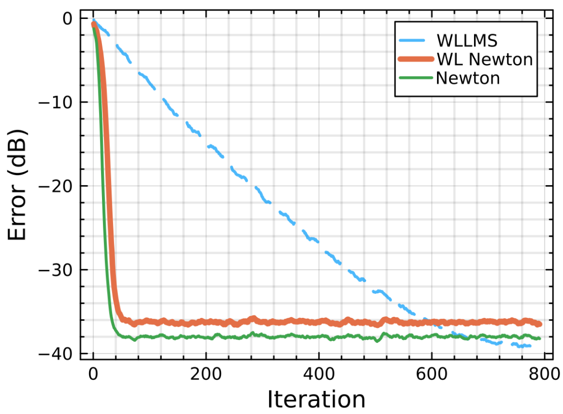


Fig. 23. Adaptive signal estimation using the Newton algorithms for a complex non-circular Gaussian signal with non-circular noise.

Fig. 24 depicts the same process, using the desired signal as the real part, following the methodology applied in previous simulations. In this scenario, there is a slight improvement in

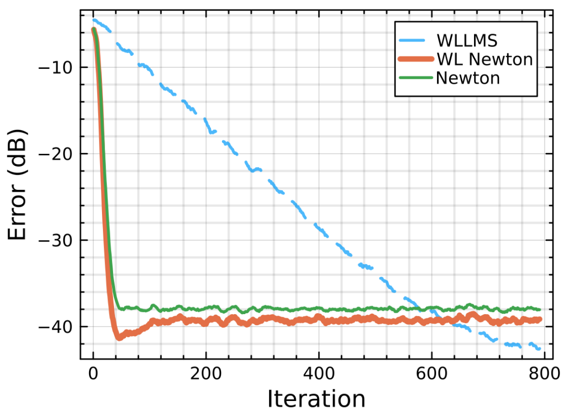


Fig. 24. Adaptive signal estimation using the Newton algorithms for the real part of a complex non-circular Gaussian signal with non-circular noise.

the performance of the augmented algorithm. In both cases, the

Hessians-weighted algorithms are compared with the WLLMS algorithm to demonstrate the faster convergence rate of the Newton algorithms. The fact that the WLLMS algorithm reaches lower error levels at the end of execution is mainly attributed to the use of a step size parameter (or learning rate), which controls the rate of adaptation. In contrast, the Newton method does not inherently include such a step-size mechanism, unless explicitly modified. As a result, although Newton-based algorithms typically converge faster, they may not always reach the lowest possible error unless additional mechanisms, such as step-size tuning or regularization, are introduced.

X. Conclusion

This tutorial has been crafted with the aim of offering a thorough exploration of the theoretical foundations and some applications of adaptive filtering for complex-valued signals. The integration of complex values into signal processing has proven to be a powerful tool, simplifying calculations and expanding the scope of capabilities.

The intrinsic geometry of complex-valued data, arising from the two-component representation, introduces new considerations in signal processing. As demonstrated in this tutorial, understanding the profile of data is crucial for optimizing performance in various signal processing applications.

By adopting a special notation, namely complex augmented representation, and employing a widely linear processing structure, superior performance can be achieved in comparison to conventional complex filtering methods in specific scenarios. The purpose of the tutorial has been to demystify the challenges associated with processing complex-valued signals and to highlight the importance of considering the data profile for meaningful information extraction. The techniques presented here can be readily applied to other models or constructions of interest, providing a versatile toolkit for signal processing applications. This tutorial stands as a testament to the ongoing evolution of signal processing methodologies in the hope that it inspires further exploration and innovation in this field.

References

[1] A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[2] J. E and Z. Lu, "Least-Squares Algorithms for Complex-Valued Blind Source Separation," *Circuits, Systems, and Signal Processing*, 2024.

[3] D. B. Rowe and B. R. Logan, "A Complex Way to Compute fMRI Activation," *NeuroImage*, vol. 23, no. 3, pp. 1078–1092, 2004.

[4] D. P. Mandic, S. Javidi, V. S. L. Goh, A. Kuh, and K. Aihara, "Complex-Valued Prediction of Wind Profile Using Augmented Complex Statistics," *Renewable Energy*, vol. 34, no. 1, pp. 196–201, 2009.

[5] D. J. Griffiths and D. F. Schroeter, *Introduction to Quantum Mechanics*, 3rd ed. Cambridge: Cambridge University Press, 2018.

[6] Z. Lei, S. Gao, H. Hasegawa, Z. Zhang, M. Zhou, and K. Sedraoui, "Fully Complex-Valued Gated Recurrent Neural Network for Ultrasound Imaging," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2023.

[7] C. Lee, H. Hasegawa, and S. Gao, "Complex-Valued Neural Networks: A Comprehensive Survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 8, pp. 1406–1426, 2022.

[8] S. Mahajan and A. K. Pandit, "Hybrid Method to Supervise Feature Selection Using Signal Processing and Complex Algebra Techniques," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8213–8234, 2023.

[9] J. Xu, C. Wu, S. Ying, and H. Li, "The Performance Analysis of Complex-Valued Neural Network in Radio Signal Recognition," *IEEE Access*, vol. 10, pp. 48 708–48 718, 2022.

[10] A. Iqbal, M. Nait-Meziane, A.-K. Seghouane, and K. Abed-Meraim, "Adaptive Complex-Valued Dictionary Learning: Application to fMRI Data Analysis," *Signal Processing*, vol. 166, p. 107263, 2020.

[11] T. Zhang, W. Chen, and M. Li, "Complex-Valued Distribution Entropy and Its Application for Seizure Detection," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 306–323, 2020.

[12] J. W. Brown and R. V. Churchill, *Complex Variables and Applications*. New York: McGraw-Hill Education, 2013.

[13] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. San Diego: Academic Press, 1985.

[14] D. J. Acheson, *Elementary Fluid Dynamics*, ser. Elementary Fluid Dynamics. Clarendon Press, 1990.

[15] S. M. Wentworth, *Fundamentals of Electromagnetics with Engineering Applications*. Wiley, 2005.

[16] R. E. Greene and S. G. Krantz, *Function Theory of One Complex Variable*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2006.

[17] L.-s. Hahn, *Complex Numbers and Geometry*. Mathematical Association of America, 1994.

[18] H. Schwerdtfeger, *Geometry of Complex Numbers: Circle Geometry, Moebius Transformation, Non-Euclidean Geometry*, ser. Dover Books on Advanced Mathematics. Dover, 1979.

[19] R. A. Wooding, "The Multivariate Distribution of Complex Normal Variables," *Biometrika*, vol. 43, no. 1-2, pp. 212–215, 1956.

[20] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, 1963.

[21] F. D. Neeser and J. L. Massey, "Proper Complex Random Processes with Applications to Information Theory," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1293–1302, 1993.

[22] B. Picinbono, "On Circularity," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.

[23] B. Picinbono and P. Chevalier, "Widely Linear Estimation with Complex Data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[24] B. Picinbono, "Second-Order Complex Random Vectors and Normal Distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2637–2640, 1996.

[25] B. Picinbono and P. Bondon, "Second-Order Statistics of Complex Signals," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 411–420, 1997.

[26] E. Ollila, "On the Circularity of a Complex Random Variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.

[27] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.

[28] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[29] T. Adalı and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2010.

[30] J. Eriksson, E. Ollila, and V. Koivunen, "Essential Statistics and Tools for Complex Random Variables," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5400–5408, 2010.

[31] P. J. Schreier, L. L. Scharf, and A. Hanssen, "A Generalized Likelihood Ratio Test for Impropriety of Complex Signals," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 433–436, 2006.

[32] B. Jelfs, D. P. Mandic, and S. C. Douglas, "An Adaptive Approach for the Identification of Improper Complex Signals," *Signal Processing*, vol. 92, no. 2, pp. 335–344, 2012.

[33] T. Adalı and P. J. Schreier, "Optimization and Estimation of Complex-Valued Signals: Theory and Applications in Filtering and Blind Source Separation," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 112–128, 2014.

[34] E. C. Mengüç and N. Acır, "An Augmented Complex-Valued Least-Mean Kurtosis Algorithm for the Filtering of Noncircular Signals," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 438–448, 2018.

[35] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, 2nd ed. New York: Pearson, 1996.

[36] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Englewood Cliffs: Prentice-Hall, 1999.

[37] A. D. Poularikas, *Discrete Random Signal Processing and Filtering Primer with MATLAB*. Boca Raton, FL: CRC Press, 2008.

[38] P. J. Antsaklis and A. N. Michel, *Linear Systems*. New York: McGraw-Hill, 1997.
[39] B. A. Shenoi, *Introduction to Digital Signal Processing and Filter Design*. New York: John Wiley & Sons, Ltd, 2005.
[40] N. Benvenuto and G. Cherubini, *Adaptive Transversal Filters*. John Wiley & Sons, Ltd, 2002, ch. 3, pp. 165–254.
[41] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
[42] S. Elaydi, *An Introduction to Difference Equations*, ser. Undergraduate Texts in Mathematics. Springer New York, 2005.
[43] M. Bellanger, *Digital Signal Processing: Theory and Practice*, 10th ed. Wiley, 2024.
[44] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. Pearson, 2006.
[45] S. Haykin, *Adaptive Filter Theory*, 3rd ed. USA: Prentice-Hall, Inc., 1996.
[46] X.-L. Li and T. Adalı, "Complex-Valued Linear and Widely Linear Filtering Using MSE and Gaussian Entropy," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5672–5684, 2012.
[47] L. Hörmander, *An Introduction to Complex Analysis in Several Variables*, ser. North-Holland Mathematical Library. North-Holland, 1973.
[48] K. Kreutz-Delgado, "The Complex Gradient Operator and the CR-Calculus," 2009. [Online]. Available: https://arxiv.org/abs/0906.4835
[49] T. Adalı, H. Li, and R. Aloysius, "On Properties of the Widely Linear MSE Filter and Its LMS Implementation," in *2009 43rd Annual Conference on Information Sciences and Systems*, 2009, pp. 876–881.
[50] R. C. Bluhm and C. C. Cavalcante, "Wirtinger Calculus and Complex Natural Gradient Algorithm," *XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2022.
[51] A. van den Bos, "Complex Gradient and Hessian," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 141, no. 6, pp. 380–382, 1994.
[52] H. Li and T. Adalı, "Complex-Valued Adaptive Signal Processing Using Nonlinear Functions," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 765615, 2008.
[53] S. Axler, *Linear Algebra Done Right*, 4th ed. Springer, 2023.
[54] A. R. Hadd and J. L. Rodgers, *Understanding Correlation Matrices*, 1st ed. United States: SAGE Publications, Inc, 2021.
[55] P. Walters, *An Introduction to Ergodic Theory*, ser. Graduate Texts in Mathematics. Springer New York, 2000.
[56] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA: MIT Press, 1949.
[57] T. Adalı, P. J. Schreier, and L. L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, 2011.
[58] A. K. Boyat and B. K. Joshi, "A Review Paper: Noise Models in Digital Image Processing," 2015. [Online]. Available: https://arxiv.org/abs/1505.03489
[59] P. J. Schreier and L. L. Scharf, "Second-Order Analysis of Improper Complex Random Vectors and Processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.
[60] K. I. Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer Cham, 2018.
[61] J. Eriksson, E. Ollila, and V. Koivunen, "Statistics for Complex Random Variables Revisited," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 3565–3568.
[62] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for Complex Variables and Signals — Part I: Variables," *Signal Processing*, vol. 53, no. 1, pp. 1–13, 1996.
[63] A. Gut, *An Intermediate Course in Probability*, ser. Springer Texts in Statistics. Springer New York, 2009.
[64] K. Pearson, "Mathematical Contributions to the Theory of Evolution.—III. Regression, Heredity, and Panmixia," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 187, pp. 253–318, 1896.
[65] I. Cohen, Y. Huang, J. Chen, and J. Benesty, "Pearson Correlation Coefficient," *Noise Reduction in Speech Processing*, pp. 1–4, 2009.
[66] J. L. Rodgers and W. A. Nicewander, "Thirteen Ways to Look at the Correlation Coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.
[67] O. J. Dunn and V. A. Clark, *Applied Statistics: Analysis of Variance and Regression*, 2nd ed. New York: Wiley, 1987.
[68] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, ser. Frontiers in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics, 1995.
[69] ——, *Iterative Methods for Optimization*. Society for Industrial and Applied Mathematics, 1999.
[70] B. Widrow, J. McCool, and M. Ball, "The Complex LMS Algorithm," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 719–720, 1975.
[71] D. H. Brandwood, "A Complex Gradient Operator and Its Application in Adaptive Array Theory," *IEE Proceedings H Microwaves, Optics and Antennas*, vol. 130, no. 1, pp. 11–16, 1983.
[72] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman, "Julia: A Fast Dynamic Language for Technical Computing," 2012. [Online]. Available: https://arxiv.org/abs/1209.5145
[73] E. Matthes, *Python Crash Course: A Hands-On, Project-Based Introduction to Programming*, 2nd ed. San Francisco: No Starch Press, 2019.
[74] D. J. Tylavsky and G. R. L. Sohie, "Generalization of the Matrix Inversion Lemma," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1050–1052, 1986.

Rafael de Carvalho Bluhm received the B.Sc. degree in Teleinformatics Engineering from the Federal University of Ceará (UFC), Brazil, in 2019. During his undergraduate studies, he developed research projects on heterogeneous communication networks, distributed information systems with complex signals, and statistical signal processing, with a focus on adaptive filtering and differential geometry methods. His undergraduate thesis was entitled "Non-Circular Complex Signals: Statistical Analysis and Adaptive Filtering," under the supervision of Prof. Charles Casimiro Cavalcante, supported by CNPq. He received the M.Sc. degree in Teleinformatics Engineering from UFC in 2023, with the dissertation entitled "Improper Complex Signals: Adaptive Filtering and Natural Gradient," also under the supervision of Prof. Charles Casimiro Cavalcante, supported by CAPES. Since 2023, he has been pursuing the Ph.D. degree in Teleinformatics Engineering at UFC, under the supervision of Prof. Charles Casimiro Cavalcante, with research focused on topological signal processing and data analysis, supported by a CNPq doctoral fellowship. His research interests include statistical signal processing, adaptive filtering, complex and improper signals, differential geometry methods for signal processing, and emerging approaches in topological data analysis applied to communications.

João César Moura Mota Bachelor's degree in Physics from the Federal University of Ceará in July 1978, Master's degree in Electrical Engineering from PUC/Rio in September 1984, and PhD in Electrical Engineering from UNICAMP in August 1992. He has been a Professor at the Federal University of Ceará since 1979, where he served as Full Professor since 1999, and was granted the title of Emeritus Professor of UFC in 2024. He founded the Department and the Graduate Program (Master's and PhD) in Teleinformatics Engineering, serving as its Head and Coordinator, respectively. He conceived and founded the undergraduate programs in Teleinformatics Engineering and Telecommunications Engineering. He taught various curricular components in the area of Telecommunications Systems at the undergraduate, strict sensu graduate, and specialization levels. As a member of the Departments of Electrical Engineering and Teleinformatics Engineering, he developed R&D activities in several areas, including signal processing in digital communication systems, and created the field of Educometry, with numerous supervised master's theses and doctoral dissertations—some under co-supervision arrangements—and more than 300 papers published in national and international congresses and journals, for which he served on editorial boards. He worked as a doctoral intern at CPqD from 1987 to 1989 and at the Centre National des Télécommunications (CNET) of France Télécom in Lannion, France, from 1989 to 1990. He participated in international collaborations, coordinating European projects such as ALFA, ALBAN, and EUBRANEX, as well as Franco-Brazilian initiatives like CAPES-COFECUB and BRAFITEC/CAPES/CDEFI. He was responsible for creating the first Undergraduate Dual-Degree Program in Brazil in partnership with France. He has been a CNPq productivity fellow for several years, a CNPq post-doctoral fellow, and a visiting professor at the Institut National des Télécommunications (INT) in Evry, France, from August 1996 to July 1998. He served as a visiting professor at the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN) in France from June to July 2006, and took a one-year sabbatical at the Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis (I3S/CNRS), France, from August 2008 to July 2009, funded by a CAPES grant. At UFC, he was also responsible for developing and leading various scientific and academic projects, contracts, and agreements in collaborations with national and international academic institutions and companies, such as ERICSSON, TELECEARÁ, the German Aerospace Center (in German: Deutsches Zentrum für Luft- und Raumfahrt – DLR), ITA, AEB, INPE, and various universities and engineering schools in France, including Télécom Paris, Télécom Saint Etienne, Télécom SudParis, Écoles Centrales, Institut National des Sciences Appliquées – INSA, École Supérieure d'Électricité – SUPELEC, Institut National des Télécommunications, I3S/CNRS, Université Côte d'Azur – France, among others. He was Director of the Integrated School of Academic Development and Innovation – EIDEIA at UFC, and is a professor in the Graduate Program in Teleinformatics Engineering at UFC. Professor João Cesar has always been actively involved in the Brazilian Telecommunications Society, of which he was a Founding Member and a member of its Board.

Charles Casimiro Cavalcante received the B.Sc. degree in Electrical Engineering and the M.Sc. degree in Electrical Engineering from the Federal University of Ceará (UFC), Brazil, in 1999 and 2001, respectively. He earned the Ph.D. degree in Electrical Engineering from the State University of Campinas (UNICAMP), Brazil, in 2004. From 2004 to 2007, he was a CNPq/DCR researcher affiliated with the Department of Teleinformatics Engineering (DETI) at UFC, where he served as Visiting Professor from 2007 to 2008. Since 2008, he has been a full professor at DETI-UFC, currently holding the position of Full Professor and permanent member of the Graduate Program in Teleinformatics Engineering (PPGETI). He coordinated PPGETI from 2008 to 2012 and served as vice-coordinator from 2012 to 2018. He was a visiting professor at the Computer Science and Electrical Engineering Department, University of Maryland, Baltimore County (UMBC), USA, from 2014 to 2015, and a senior visiting professor at Laboratoire Lagrange, Université Côte d'Azur, France, from 2023 to 2024. His research focuses on signal processing applications in communications and physical layer optimization of wireless communication systems. He co-authored the books Unsupervised Signal Processing: Channel Equalization and Source Separation (CRC Press) and Machine Learning for Spectrum Sharing: A Survey (Now Publishers), and co-edited Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing (CRC Press). He is a Senior Member of the IEEE and a Senior Member of the Brazilian Telecommunications Society (SBrT), having held various leadership positions including President (2018–2020). He served as Co-Editor-in-Chief of the Journal of Communication and Information Systems (JCIS) from 2014 to 2018. Since 2018, he has been the Deputy Coordinator of Professional Programs in Engineering IV at CAPES and a full member of the Electrical and Biomedical Engineering Advisory Committee (CA-EE) at CNPq since 2024, representing Telecommunications Systems until 2027. He also served on the IEEE Signal Processing Society Board of Governors as Regional Director-at-Large for Regions 7 and 9 (2020–2021).