Complex Valued Nonlinear Adaptive Filters

Noncircularity, Widely Linear and Neural Models

Danilo P. Mandic

Imperial College London, UK

Vanessa Su Lee Goh

Shell EP, Europe

WILEY

A John Wiley and Sons, Ltd, Publication

Complex Valued Nonlinear Adaptive Filters

Complex Valued Nonlinear Adaptive Filters

Noncircularity, Widely Linear and Neural Models

Danilo P. Mandic

Imperial College London, UK

Vanessa Su Lee Goh

Shell EP, Europe

WILEY

A John Wiley and Sons, Ltd, Publication

This edition first published 2009

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Mandic, Danilo P.
Complex valued nonlinear adaptive filters : noncircularity, widely linear, and neural models / by Danilo P. Mandic, Vanessa Su Lee Goh, Shell EP, Europe.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-06635-5 (cloth)
1. Functions of complex variables. 2. Adaptive filters–Mathematical models. 3. Filters (Mathematics) 4. Nonlinear theories. 5. Neural networks (Computer science) I. Goh, Vanessa Su Lee. II. Holland, Shell. III. Title.
TA347.C64.M36 2009
621.382'2–dc22 2009001965

A catalogue record for this book is available from the British Library.

ISBN: 978-0-470-06635-5

Typeset in 10/12 pt Times by Thomson Digital, Noida, India

Printed in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire

The real voyage of discovery consists
not in seeking new landscapes
but in having new eyes
Marcel Proust

Adaptive and Learning Systems for Signal, Processing, Communications, and Control

Editor: Simon Haykin

Beckerman / ADAPTIVE COOPERATIVE SYSTEMS
Candy / MODEL-BASED SIGNAL PROCESSING
Chen, Haykin, Eggermont, and Becker / CORRELATIVE LEARNING: A Basis for Brain and Adaptive Systems
Chen and Gu / CONTROL-ORIENTED SYSTEM IDENTIFICATION: An H_∞ Approach
Cherkassky and Mulier / LEARNING FROM DATA: Concepts, Theory, and Methods
Diamantaras and Kung / PRINCIPAL COMPONENT NEURAL NETWORKS: Theory and Applications
Haykin / UNSUPERVISED ADAPTIVE FILTERING: Blind Source Separation
Haykin / UNSUPERVISED ADAPTIVE FILTERING: Blind Deconvolution
Haykin and Puthussarypady / CHAOTIC DYNAMICS OF SEA CLUTTER
Haykin and Widrow / LEAST MEAN-SQUARE ADAPTIVE FILTERS
Hrycej / Neurocontrol: TOWARDS AN INDUSTRIAL CONTROL METHODOLOGY
Hyvärinen, Karhunen, and Oja / INDEPENDENT COMPONENT ANALYSIS
Kristić, Kanellakopoulos, and Kokotović / NONLINEAR AND ADAPTIVE CONTROL DESIGN
Mandic and Goh / COMPLEX VALUED NONLINEAR ADAPTIVE FILTERS: Non Circularity, widely Linear and Neural Models
Mann / Intelligent Image Processing
Nikias and Shao / SIGNAL PROCESSING WITH ALPHA-STABLE DISTRIBUTIONS AND APPLICATIONS
Passino and Burgess / STABILITY ANALYSIS OF DISCRETE EVENT SYSTEMS
Sánchez-Peña and Sznaier / ROBUST SYSTEMS THEORY AND APPLICATIONS
Sandberg, Lo, F ancourt, Principe, Katagairi, and Haykin / NONLINEAR DYNAMICAL SYSTEMS: Feedforward Neural Network Perspectives
Spooner, Maggiore, Ordóñez, and Passino / STABLE ADAPTIVE CONTROL AND ESTIMATION FOR NONLINEAR SYSTEMS: Neural and Fuzzy Approximator Techniques
Tao / ADAPTIVE CONTROL DESIGN AND ANALYSIS
Tao and Kokotović / ADAPTIVE CONTROL OF SYSTEMS WITH ACTUATOR AND SENSOR NONLINEARITIES
Tsoukalas and Uhrig / FUZZY AND NEURAL APPROACHES IN ENGINEERING
Van Hulle / FAITHFUL REPRESENTATIONS AND TOPOGRAPHIC MAPS: From Distortion- to Information-Based Self-Organization
Vapnik / STATISTICAL LEARNING THEORY
Werbos / THE ROOTS OF BACKPROPAGATION: From Ordered Derivatives to Neural Networks and Political Forecasting
Yee and Haykin / REGULARIZED RADIAL BIAS FUNCTION NETWORKS: Theory and Applications

Contents

Preface

This book was written in response to the growing demand for a text that provides a unified treatment of complex valued adaptive filters, both linear and nonlinear, and methods for the processing of both complex circular and complex noncircular signals. We believe that this is the first attempt to bring together established adaptive filtering algorithms in $\mathbb{C}$ and the recent developments in the statistics of complex variable under the umbrella of powerful mathematical frameworks of $\mathbb{CR}$ (Wirtinger) calculus and augmented complex statistics. Combining the results from the authors' original research and current established methods, this books serves as a rigorous account of existing and novel complex signal processing methods, and provides next generation solutions for adaptive filtering of the generality of complex valued signals. The introductory chapters can be used as a text for a course on adaptive filtering. It is our hope that people as excited as we are by the possibilities opened by the more advanced work in this book will further develop these ideas into new and useful applications.

The title reflects our ambition to write a book which addresses several major problems in modern complex adaptive filtering. Real world data are non-Gaussian, nonstationary and generated by nonlinear systems with possibly long impulse responses. For the processing of such signals we therefore need *nonlinear* architectures to deal with nonlinearity and non-Gaussianity, *feedback* to deal with long responses, and *adaptive* mode of operation to deal with the nonstationary nature of the data. These have all been brought together in this book, hence the title "*Complex Valued Nonlinear Adaptive Filters*". The subtitle reflects some more intricate aspects of the processing of complex random variables, and that the class of nonlinear filters addressed in this work can be viewed as temporal neural networks. This material can also be used to supplement courses on neural networks, as the algorithms developed can be used to train neural networks for pattern processing and classification.

Complex valued signals play a pivotal role in communications, array signal processing, power, environmental, and biomedical signal processing and related fields. These signals are either complex by design, such as symbols used in data communications (e.g. quadrature phase shift keying), or are made complex by convenience of representation. The latter class includes analytic signals and signals coming from many important modern applications in magnetic source imaging, interferometric radar, direction of arrival estimation and smart antennas, mathematical biosciences, mobile communications, optics and seismics. Existing books do not take into account the effects on performance of a unique property of complex statistics – complex noncircularity, and employ several convenient mathematical shortcuts in the treatment of complex random variables.

Adaptive filters based on widely linear models introduced in this work are derived rigorously, and are suited for the processing of a much wider class of *complex noncircular signals* (directional processes, vector fields), and offer a number of theoretical performance gains.

Perhaps the first time we became involved in practical applications of complex adaptive filtering was when trying to perform short term wind forecasting by treating wind speed and direction, which are routinely processed separately, as a unique complex valued quantity. Our results outperformed the standard approaches. This opened a can of worms, as it became apparent that the standard techniques were not adequate, and that mathematical foundations and practical tools for the applications of complex valued adaptive filters to the generality of complex signals are scattered throughout the literature. For instance, an often confusing aspect of complex adaptive filtering is that the cost (objective) function to be minimised is a real function (measure of error power) of complex variables, and is not analytic. Thus, standard complex differentiability (Cauchy-Riemann conditions) does not apply, and we need to resort to pseudoderivatives. We identified the need for a rigorous, concise, and unified treatment of the statistics of complex variables, methods for dealing with nonlinearity and noncircularity, and enhanced solutions for adaptive signal processing in $\mathbb{C}$, and were encouraged by our series editor Simon Haykin and the staff from Wiley Chichester to produce this text.

The first two chapters give the introduction to the field and illustrate the benefits of the processing in the complex domain. Chapter 1 provides a personal view of the history of complex numbers. They are truly fascinating and, unlike other number systems which were introduced as solutions to practical problems, they arose as a product of intellectual exercise. Complex numbers were formalised in the mid-19th century by Gauss and Euler in order to provide solutions for the fundamental theorem of algebra; within 50 years (and without the Internet) they became a linchpin of electromagnetic field and relativity theory. Chapter 2 offers theoretical and practical justification for converting many apparently real valued signal processing problems into the complex domain, where they can benefit from the convenience of representation and the power and beauty of complex calculus. It illustrates the duality between the processing in $\mathbb{R}^2$ and $\mathbb{C}$, and the benefits of complex valued processing – unlike $\mathbb{R}^2$ the field of complex numbers forms a division algebra and provides a rigorous mathematics framework for the treatment of phase, nonlinearity and coupling between signal components.

The foundations of standard complex adaptive filtering are established in Chapters 3–7. Chapter 3 provides an overview of adaptive filtering architectures, and introduces the background for their state space representations and links with polynomial filters and neural networks. Chapter 4 deals with the choice of complex nonlinear activation function and addresses the trade off between their boundedness and analyticity. The only continuously differentiable function in $\mathbb{C}$ that satisfies the Cauchy-Riemann conditions is a constant; to preserve boundedness some ad hoc approaches (also called split-complex) employ real valued nonlinearities on the real and imaginary parts. Our main interest is in complex functions of complex variables (also called fully complex) which are not bounded on the whole complex plane, but are complex differentiable and provide solutions which are generic extensions of the corresponding solutions in $\mathbb{R}$. Chapter 5 addresses the duality between gradient calculation in $\mathbb{R}^2$ and $\mathbb{C}$ and introduces the so called $\mathbb{CR}$ calculus which is suitable for general functions of complex variables, both holomorphic and non-holomorphic. This provides a unified framework for computing the Jacobians, Hessians, and gradients of cost functions, and serves as a basis for the derivation of learning algorithms throughout this book. Chapters 6 and 7 introduce standard complex valued adaptive filters, both linear and nonlinear; they are supported by rigorous proofs of convergence, and can be used to teach a course on adaptive filtering. The complex least mean square (CLMS) in Chapter 6 is derived step by step, whereas the learning algorithms for feedback structures in Chapter 7 are derived in a compact way, based on $\mathbb{CR}$

calculus. Furthermore, learning algorithms for both linear and nonlinear feedback architectures are introduced, starting from linear IIR filters to temporal recurrent neural networks.

Chapters 8–11 address several practical aspects of adaptive filtering, such as adaptive step-sizes, dynamical range extension, and a posteriori mode of operation. Chapter 8 provides a thorough review of adaptive step size algorithms and introduces the general normalised gradient descent (GNGD) algorithm for enhanced stability. Chapter 9 gives solutions for dynamical range extension of nonlinear neural adaptive filters, whereas Chapter 10 explains a posteriori algorithms and analyses them in the framework of fixed point theory. Chapter 11 rounds up the first part of the book and introduces fractional delay filters together with links between complex nonlinear functions and number theory.

Chapters 12–15 introduce linear and nonlinear adaptive filters based on widely linear models, which are suited to deal with complex noncircularity, thus providing theoretical and practical adaptive filtering solutions for the generality of complex signals. Chapter 12 provides a comprehensive overview of the latest results (2008) in the statistics of complex random signals, with a particular emphasis on complex noncircularity. It is shown that the standard complex Gaussian model is inadequate and the concepts of noise, stationarity, multicorrelation, and multispectra are re-introduced based on the augmented statistics. This has served as a basis for the development of the class of ‘augmented’ adaptive filtering algorithms, starting from the complex least square (ACLMS) algorithm through to augmented learning algorithms for IIR filters, recurrent neural networks, and augmented Kalman filters. Chapter 13 introduces the augmented least mean square algorithm, a quantum step in the adaptive signal processing of complex noncircular signals. It is shown that this approach is as good as standard approaches for circular data, whereas it outperforms standard filters for noncircular data. Chapter 14 provides an insight into the duality between complex valued linear adaptive filters and dual channel real adaptive filters. A correspondence is established between the ACLMS and the dual channel real LMS algorithms. Chapter 15 extends widely linear modelling in $\mathbb{C}$ to feedback and nonlinear architectures. The derivations are based on $\mathbb{CR}$ calculus and are provided for both the gradient descent and state space (Kalman filtering) models.

Chapter 16 addresses collaborative adaptive filtering in $\mathbb{C}$. It is shown that by employing collaborative filtering architectures we can gain insight into the nature of a signal in hand, and a simple test for complex noncircularity is proposed. Chapter 17 introduces complex empirical mode decomposition (EMD), a data driven time-frequency technique. This technique, when used for preprocessing complex valued data, provides a framework for “data fusion via fission”, with a number of applications, especially in biomedical engineering and neuroscience. Chapter 18 provides a rigorous statistical testing framework for the validity of complex representation.

The material is supported by a number of Appendices (some of them based on [190]), ranging from the theory of complex variable through to fixed point theory. We believe this makes the book self-sufficient for a reader who has basic knowledge of adaptive signal processing. Simulations were performed for both circular and noncircular data sources, from benchmark linear and nonlinear models to real world wind and radar signals. The applications are set in a prediction setting, as prediction is at the core of adaptive filtering. The complex valued wind signal is our most frequently used test signal, due to its intermittent, non-Gaussian and noncircular nature. Gill Instruments provided ultrasonic anemometers used for our wind recordings.

Acknowledgements

Vanessa and I would like to thank our series editor Simon Haykin for encouraging us to write a text on modern complex valued adaptive signal processing. In addition, my own work in this area was inspired by the success of my earlier monograph "*Recurrent Neural Networks for Prediction*", Wiley 2001, co-authored with Jonathon Chambers, where some earlier results were outlined. Over the last seven years these ideas have matured greatly, through working with my co-author Vanessa Su Lee Goh and a number of graduate students, to a point where it was possible to write this book. I have had great pleasure to work with Temujin Gautama, Maciej Pedzisz, Mo Chen, David Looney, Phebe Vayanos, Beth Jelfs, Clive Cheong Took, Yili Xia, Andrew Hanna, Christos Boukis, George Souretis, Naveed Ur Rehman, Tomasz Rutkowski, Toshihisa Tanaka, and Soroush Javidi (who has also designed the book cover), who have all been involved in the research that led to this book. Their dedication and excitement have helped to make this journey through the largely unchartered territory of modern complex valued signal processing so much more rewarding.

Peter Schreier has provided deep and insightful feedback on several chapters, especially when it comes to dealing with complex noncircularity. We have enjoyed the interaction with Tülay Adalı, who also proofread several key chapters. Ideas on the duality between real and complex filters matured through discussions with Susanna Still and Jacob Benesty. The collaboration with Scott Douglas influenced convergence proofs in Chapter 6. The results in Chapter 18 arose from collaboration with Marc Van Hulle and his team. Tony Constantinides, Igor Aizenberg, Aurelio Uncini, Tony Kuh, Preben Kidmose, Maria Petrou, Isao Yamada, and Olga Boric Lubecke provided valuable comments.

Additionally, I would like to thank Andrzej Cichocki for invigorating discussions and the timely reminder that the quantum developments of science are in the hands of young researchers. Consequently, we decided to hurry up with this book while I can still (just) qualify. The collaboration with Kazuyuki Aihara and Yoshito Hirata helped us to hone our ideas related to complex valued wind forecasting.

It is not possible to mention all the colleagues and friends who have helped towards this book. Members of the IEEE Signal Processing Society Technical Committee on Machine Learning for Signal Processing have provided support and stimulating discussions, in particular, David Miller, Dragan Obradovic, Jose Principe, and Jan Larsen. We wish to express our appreciation to the signal processing tradition and vibrant research atmosphere at Imperial College London, which have made delving into this area so rewarding.

We are deeply indebted to Henry Goldstein, who tamed our immense enthusiasm for the subject and focused it to the needs of our readers.

Finally, our love and gratitude goes to our families and friends for supporting us since the summer of 2006, when this work began.

Danilo P. Mandic
Vanessa Su Lee Goh

1

The Magic of Complex Numbers

The notion of complex number is intimately related to the *Fundamental Theorem of Algebra* and is therefore at the very foundation of mathematical analysis. The development of complex algebra, however, has been far from straightforward.[1]

The human idea of 'number' has evolved together with human society. The *natural* numbers ($1, 2, \ldots \in \mathbb{N}$) are straightforward to accept, and they have been used for *counting* in many cultures, irrespective of the actual base of the number system used. At a later stage, for *sharing*, people introduced fractions in order to answer a simple problem such as 'if we catch $\mathcal{U}$ fish, I will have two parts $\frac{2}{5}\,\mathcal{U}$ and you will have three parts $\frac{3}{5}\,\mathcal{U}$ of the whole catch'. The acceptance of negative numbers and zero has been motivated by the emergence of economy, for dealing with profit and loss. It is rather impressive that ancient civilisations were aware of the need for irrational numbers such as $\sqrt{2}$ in the case of the Babylonians [77] and π in the case of the ancient Greeks.[2]

The concept of a new 'number' often came from the need to solve a specific practical problem. For instance, in the above example of sharing $\mathcal{U}$ number of fish caught, we need to solve for $2\mathcal{U} = 5$ and hence to introduce fractions, whereas to solve $x^2 = 2$ (diagonal of a square) irrational numbers needed to be introduced. Complex numbers came from the necessity to solve equations such as $x^2 = -1$.

[1] A classic reference which provides a comprehensive account of the development of numbers is *Number: The Language of Science* by Tobias Dantzig [57].

[2] The Babylonians have actually left us the basics of Fixed Point Theory (see Appendix P), which in terms of modern mathematics was introduced by Stefan Banach in 1922. On a clay tablet (YBC 7289) from the Yale Babylonian Collection, the Mesopotamian scribes explain how to calculate the diagonal of a square with base 30. This was achieved using a fixed point iteration around the initial guess. The ancient Greeks used π in geometry, although the irrationality of π was only proved in the 1700s. More information on the history of mathematics can be found in [34] whereas P. Nahin's book is dedicated to the history of complex numbers [215].

Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models
Danilo P. Mandic and Vanessa Su Lee Goh

1.1 History of Complex Numbers

Perhaps the earliest reference to square roots of negative numbers occurred in the work of Heron of Alexandria[3], around 60 AD, who encountered them while calculating volumes of geometric bodies. Some 200 years later, Diophantus (about 275 AD) posed a simple problem in geometry,

Find the sides of a right–angled triangle of perimeter 12 units and area 7 squared units.

which is illustrated in Figure 1.1. To solve this, let the side $|AB| = x$, and the height $|BC| = h$, to yield

$$area = \frac{1}{2}\,x\,h$$

$$perimeter = x + h + \sqrt{x^2 + h^2}$$

In order to solve for x we need to find the roots of

$$6x^2 - 43x + 84 = 0$$

however this equation does not have real roots.

A similar problem was posed by Cardan[4] in 1545. He attempted to find two numbers a and b such that

$$a + b = 10$$

$$a\,b = 40$$

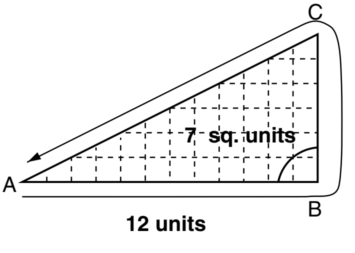


Figure 1.1 Problem posed by Diophantus (third century AD)

[3]Heron (or Hero) of Alexandria was a Greek mathematician and inventor. He is credited with finding a formula for the area of a triangle (as a function of the perimeter). He invented many gadgets operated by fluids; these include a fountain, fire engine and siphons. The aeolipile, his engine in which the recoil of steam revolves a ball or a wheel, is the forerunner of the steam engine (and the jet engine). In his method for approximating the square root of a number he effectively found a way round the complex number. It is fascinating to realise that complex numbers have been used, implicitly, long before their introduction in the 16th century.

[4]Girolamo or Hieronimo Cardano (1501–1576). His name in Latin was Hieronymus Cardanus and he is also known by the English version of his name Jerome Cardan. For more detail on Cardano's life, see [1].

These equations are satisfied for

$$a = 5 + \sqrt{-15} \quad \text{and} \quad b = 5 - \sqrt{-15} \tag{1.1}$$

which are clearly not real.

The need to introduce the complex number became rather urgent in the 16th century. Several mathematicians were working on what is today known as the *Fundamental Theorem of Algebra* (FTA) which states that

Every nth order polynomial with real[5] *coefficients has exactly n roots in* $\mathbb{C}$.

Earlier attempts to find the roots of an arbitrary polynomial include the work by Al-Khwarizmi (ca 800 AD), which only allowed for positive roots, hence being only a special case of FTA. In the 16th century Niccolo Tartaglia[6] and Girolamo Cardano (see Equation 1.1) considered closed formulas for the roots of third- and fourth-order polynomials. Girolamo Cardano first introduced complex numbers in his *Ars Magna* in 1545 as a tool for finding *real* roots of the 'depressed' cubic equation $x^3 + ax + b = 0$. He needed this result to provide algebraic solutions to the general cubic equation

$$ay^3 + by^2 + cy + d = 0$$

By substituting $y = x - \frac{1}{3}b$, the cubic equation is transformed into a depressed cubic (without the square term), given by

$$x^3 + \beta x + \gamma = 0$$

Scipione del Ferro of Bologna and Tartaglia showed that the depressed cubic can be solved as[7]

$$x = \sqrt[3]{-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} + \sqrt[3]{-\frac{\gamma}{2} - \sqrt{\frac{\gamma^2}{4} + \frac{\beta^3}{27}}} \tag{1.2}$$

For certain problem settings (for instance $a = 1, b = 9, c = 24, d = 20$), and using the substitution $y = x - 3$, Tartaglia could show that, by symmetry, there exists $\sqrt{-1}$ which has mathematical meaning. For example, Tartaglia's formula for the roots of $x^3 - x = 0$ is given by

$$\frac{1}{\sqrt{3}}\left((\sqrt{-1})^{\frac{1}{3}} + \frac{1}{(\sqrt{-1})^{\frac{1}{3}}}\right)$$

[5] In fact, it states that every nth order polynomial with complex coefficients has n roots in $\mathbb{C}$, but for historical reasons we adopt the above variant.

[6] Real name Niccolo Fontana, who is known as Tartaglia (the stammerer) due to a speaking disorder.

[7] In modern notation this can be written as $x = (q + w)^{\frac{1}{3}} + (q - w)^{\frac{1}{3}}$.

Rafael Bombelli also analysed the roots of cubic polynomials by the 'depressed cubic' transformations and by applying the Ferro–Tartaglia formula (1.2). While solving for the roots of

$$x^3 - 15x - 4 = 0$$

he was able to show that

$$\left(2 + \sqrt{-1}\right) + \left(2 - \sqrt{-1}\right) = 4$$

Indeed $x = 4$ is a correct solution, however, in order to solve for the real roots, it was necessary to perform calculations in $\mathbb{C}$. In 1572, in his *Algebra*, Bombelli introduced the symbol $\sqrt{-1}$ and established rules for manipulating 'complex numbers'.

The term 'imaginary' number was coined by Descartes in the 1630s to reflect his observation that '*For every equation of degree n, we can imagine n roots which do not correspond to any real quantity*'. In 1629, Flemish mathematician[8] Albert Girard in his *L'Invention Nouvelle en l'Algèbre* asserts that there are n roots to an nth order polynomial, however this was accepted as self-evident, but with no guarantee that the actual solution has the form $a + \jmath b$, $a, b \in \mathbb{R}$.

It was only after their geometric representation (John Wallis[9] in 1685 in *De Algebra Tractatus* and Caspar Wessel[10] in 1797 in the *Proceedings of the Copenhagen Academy*) that the complex numbers were finally accepted. In 1673, while investigating geometric representations of the roots of polynomials, John Wallis realised that for a general quadratic polynomial of the form

$$x^2 + 2bx + c^2 = 0$$

for which the solution is

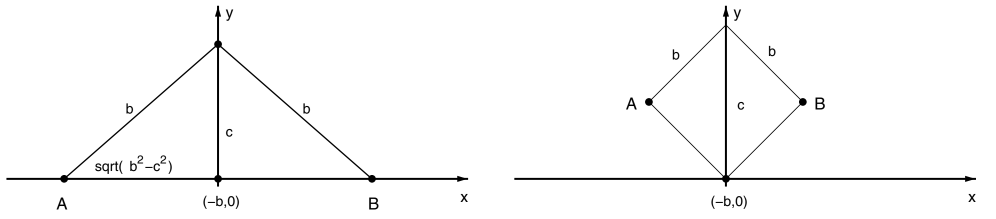
$$x = -b \pm \sqrt{b^2 - c^2} \tag{1.3}$$

a geometric interpretation was only possible for $b^2 - c^2 \geq 0$. Wallis visualised this solution as displacements from the point $-b$, as shown in Figure 1.2(a) [206]. He interpreted each solution as a vertex (A and B in Figure 1.2) of a right triangle with height c and side $\sqrt{b^2 - c^2}$. Whereas this geometric interpretation is clearly correct for $b^2 - c^2 \geq 0$, Wallis argued that for $b^2 - c^2 < 0$, since b is shorter than c, we will have the situation shown in Figure 1.2(b); this

[8] Albert Girard was born in France in 1595, but his family later moved to the Netherlands as religious refugees. He attended the University of Leiden where he studied music. Girard was the first to propose the fundamental theorem of algebra, and in 1626, in his first book on trigonometry, he introduced the abbreviations sin, cos, and tan. This book also contains the formula for the area of a spherical triangle.

[9] In his *Treatise on Algebra* Wallis accepts negative and complex roots. He also shows that equation $x^3 - 7x = 6$ has exactly three roots in $\mathbb{R}$.

[10] Within his work on geodesy Caspar Wessel (1745–1818) used complex numbers to represent directions in a plane as early as in 1787. His article from 1797 entitled 'On the Analytical Representation of Direction: An Attempt Applied Chiefly to Solving Plane and Spherical Polygons' (in Danish) is perhaps the first to contain a well-thought-out geometrical interpretation of complex numbers.



(a) Real solution (b) Complex solution

Figure 1.2 Geometric representation of the roots of a quadratic equation

way we can think of a complex number as a *point on the plane*.[11] In 1732 Leonhard Euler calculated the solutions to the equation

$$x^n - 1 = 0$$

in the form of

$$\cos\theta + \sqrt{-1}\sin\theta$$

and tried to visualise them as the vertices of a planar polygon. Further breakthroughs came with the work of Abraham de Moivre (1730) and again Euler (1748), who introduced the famous formulas

$$\begin{aligned}(\cos\theta + \jmath\sin\theta)^n &= \cos n\theta + \jmath\sin n\theta\\ \cos\theta + \jmath\sin\theta &= e^{\jmath\theta}\end{aligned}$$

Based on these results, in 1749 Euler attempted to prove FTA for real polynomials in *Recherches Sur Les Racines Imaginaires des Équations*. This was achieved based on a decomposition a monic polynomials and by using Cardano's technique from *Ars Magna* to remove the second largest degree term of a polynomial.

In 1806 the Swiss accountant and amateur mathematician Jean Robert Argand published a proof of the FTA which was based on an idea by d'Alembert from 1746. Argand's initial idea was published as *Essai Sur Une Manière de Représenter les Quantités Imaginaires Dans les Constructions Géométriques* [60, 305]. He simply interpreted $\jmath$ as a rotation by 90° and introduced the Argand plane (or Argand diagram) as a geometric representation of complex numbers. In Argand's diagram, $\pm\sqrt{-1}$ represents a unit line, perpendicular to the real axis. The notation and terminology we use today is pretty much the same. A complex number

$$z = x + \jmath y$$

[11]In his interpretation $-\sqrt{-1}$ is the same point as $\sqrt{-1}$, but nevertheless this was an important step towards the geometric representation of complex numbers.

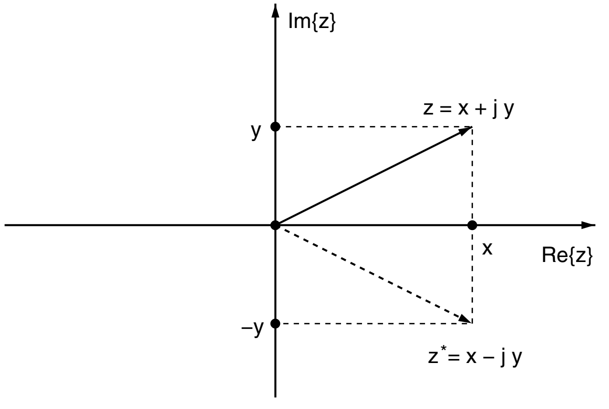


Figure 1.3 Argand's diagram for a complex number z and its conjugate z^*

is simply represented as a vector in the complex plane, as shown in Figure 1.3. Argand called $\sqrt{x^2 + y^2}$ the *modulus*, and Gauss introduced the term *complex number* and notation[12] $\imath = \sqrt{-1}$ (in signal processing we use $\jmath = \imath = \sqrt{-1}$). Karl Friedrich Gauss used complex numbers in his several proofs of the fundamental theorem of algebra, and in 1831 he not only associated the complex number $z = x + \jmath y$ with a point (x, y) on a plane, but also introduced the rules for the addition[13] and multiplication of such numbers. Much of the terminology used today comes from Gauss, Cauchy[14] who introduced the term 'conjugate', and Hankel who in 1867 introduced the term *direction coefficient* for $\cos\theta + \jmath\sin\theta$, whereas Weierstrass (1815–1897) introduced the term *absolute value* for the modulus.

Some analytical aspects of complex numbers were also developed by Georg Friedrich Bernhard Riemann (1826–1866), and those principles are nowadays the basics behind what is known as manifold signal processing.[15] To illustrate the potential of complex numbers in this context, consider the stereographic[16] projection [242] of the Riemann sphere, shown in Figure 1.4(a). In a way analogous to Cardano's 'depressed cubic', we can perform dimensionality reduction by embedding $\mathbb{C}$ in $\mathbb{R}^3$, and rewriting

$$Z = a + \jmath b, \quad (a, b, 0) \in \mathbb{R}^3$$

[12] There is a simple trap, that is, we cannot apply the identity of the type $\sqrt{ab} = \sqrt{a}\sqrt{b}$ to the 'imaginary' numbers, this would lead to the wrong conclusion $1 = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1}$, however $\sqrt{-1}^2 = \sqrt{-1}\sqrt{-1} = -1$.

[13] So much so that, for instance, 3 remains a prime number whereas 5 does not, since it can be written as $(1 - 2\jmath)(1 + 2\jmath)$.

[14] Augustin Louis Cauchy (1789–1867) formulated many of the classic theorems in complex analysis.

[15] Examples include the Natural Gradient algorithm used in blind source separation [10, 49].

[16] The stereographic projection is a mapping that projects a sphere onto a plane. The mapping is smooth, bijective and conformal (preserves relationships between angles).

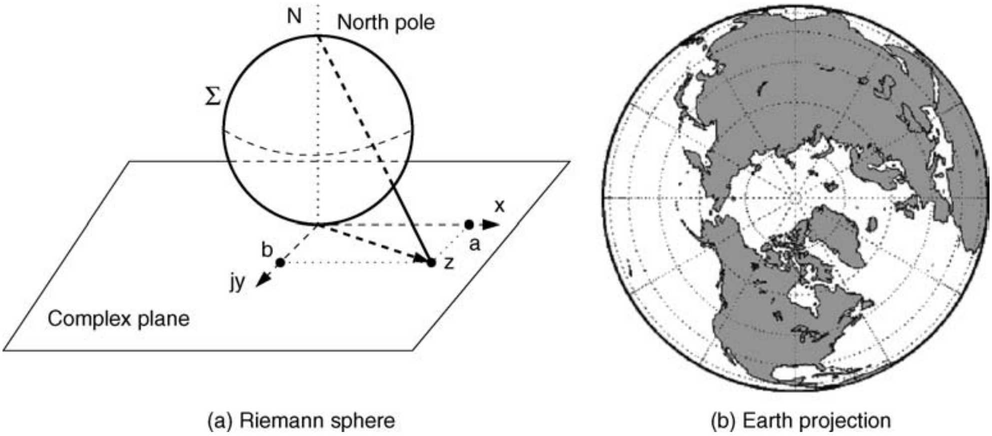


Figure 1.4 Stereographic projection and Riemann sphere: (a) the principle of the stereographic projection; (b) stereographic projection of the Earth (seen from the south pole S)

Consider a sphere Σ defined by

$$\Sigma = \left\{(x, y, u) \in \mathbb{R}^3 \,:\, x^2 + y^2 + (u - d)^2 = r^2\right\}, \quad d, r \in \mathbb{R}$$

There is a one-to-one correspondence between the points of $\mathbb{C}$ and the points of Σ, excluding N (the north pole of Σ), since the line from any point $z \in \mathbb{C}$ cuts $\Sigma \setminus \{N\}$ in precisely one point. If we include the point ∞, so as to have the *extended complex plane* $\mathbb{C} \cup \{\infty\}$, then the north pole N from sphere Σ is also included and we have a mapping of the Riemann sphere onto the extended complex plane. A stereographic projection of the Earth onto a plane tangential to the north pole N is shown in Figure 1.4(b).

1.1.1 Hypercomplex Numbers

Generalisations of complex numbers (generally termed 'hypercomplex numbers') include the work of Sir William Rowan Hamilton (1805–1865), who introduced the quaternions in 1843. A quaternion $\vec{q}$ is defined as [103]

$$\vec{q} = q_0 + q_1\imath + q_2\jmath + q_3 k \tag{1.4}$$

where the variables $\imath, \jmath, k$ are all defined as $\sqrt{-1}$, but their multiplication is not commutative.[17]

Pivotal figures in the development of the theory of complex numbers are Hermann Günther Grassmann (1809–1877), who introduced multidimensional vector calculus, and James Cockle,

[17]That is: $\imath\jmath = -\jmath\imath = k$, $\jmath k = -k\jmath = \imath$, and $k\imath = -\imath k = \jmath$.

who in 1848 introduced split-complex numbers.[18] A split-complex number (also known as motors, dual numbers, hyperbolic numbers, tessarines, and Lorenz numbers) is defined as [51]

$$z = x + \jmath y, \qquad \jmath^2 = 1$$

In 1876, in order to model spins, William Kingdon Clifford introduced a system of hypercomplex numbers (Clifford algebra). This was achieved by conveniently combining the quaternion algebra and split-complex numbers. Both Hamilton and Clifford are credited with the introduction of *biquaternions*, that is, quaternions for which the coefficients are complex numbers. A comprehensive account of *hypercomplex* numbers can be found in [143]; in general a hypercomplex number system has at least one non-real axis and is closed under addition and multiplication. Other members of the family of hypercomplex numbers include McFarlane's hyperbolic quaternion, hyper-numbers, multicomplex numbers, and twistors (developed by Roger Penrose in 1967 [233]).

1.2 History of Mathematical Notation

It is also interesting to look at the development of 'symbols' and abbreviations in mathematics. For books copied by hand the choice of mathematical symbols was not an issue, whereas for printed books this choice was largely determined by the availability of fonts of the early printers. Thus, for instance, in the 9th century in Al-Khwarizmi's *Algebra* solutions were descriptive rather than in the form of equations, while in Cardano's *Ars Magna* in the 16th century the unknowns were denoted by single roman letters to facilitate the printing process.

It was arguably Descartes who first established some general rules for the use of mathematical symbols. He used lowercase italic letters at the beginning of the alphabet to denote unknown constants (a, b, c, d), whereas letters at the end of the alphabet were used for unknown variables (x, y, z, w). Using Descartes' recommendations, the expression for a quadratic equation becomes

$$a\,x^2 + b\,x + c = 0$$

which is exactly the way we use it in modern mathematics.

As already mentioned, the symbol for imaginary unit $\imath = \sqrt{-1}$ was introduced by Gauss, whereas boldface letters for vectors were first introduced by Oliver Heaviside [115]. More details on the history of mathematical notation can be found in the two–volume book *A History of Mathematical Notations* [39], written by Florian Cajori in 1929.

In the modern era, the introduction of mathematical symbols has been closely related with the developments in computing and programming languages.[19] The relationship between computers and typography is explored in *Digital Typography* by Donald E. Knuth [153], who also developed the TeX typesetting language.

[18] Notice the difference between the split-complex *numbers* and split-complex *activation functions* of neurons [152, 190]. The term split-complex number relates to an alternative hypercomplex *number* defined by $x + \jmath y$ where $\jmath^2 = 1$, whereas the term split-complex function refers to *functions* $g : \mathbb{C} \to \mathbb{C}$ for which the real and imaginary part of the 'net' function are processed separately by a real function of real argument f, to give $g(net) = f(\Re(net)) + \jmath f(\Im(net))$.

[19] Apart from the various new symbols used, e.g. in computing, one such symbol is © for 'copyright'.

1.3 Development of Complex Valued Adaptive Signal Processing

The distinguishing characteristics of complex valued nonlinear adaptive filtering are related to the character of complex nonlinearity, the associated learning algorithms, and some recent developments in complex statistics. It is also important to notice that the universal function approximation property of some complex nonlinearities does not guarantee fast and efficient learning.

Complex nonlinearities. In 1992, Georgiou and Koutsougeras [88] proposed a list of requirements that a complex valued activation function should satisfy in order to qualify for the nonlinearity at the neuron. The calculation of complex gradients and Hessians has been detailed in work by Van Den Bos [30]. In 1995 Arena *et al.* [18] proved the *universal approximation property*[20] of a Complex Multilayer Perceptron (CMLP), based on the split-complex approach. This also gave theoretical justification for the use of complex neural networks (NNs) in time series modelling tasks, and thus gave rise to temporal neural networks. The split-complex approach has been shown to yield reasonable performance in channel equalisation applications [27, 147, 166], and in applications where there is no strong coupling between the real and imaginary part within the complex signal. However, for the common case where the inphase (I) and quadrature (Q) components have the same variance and are uncorrelated, algorithms employing split-complex activation functions tend to yield poor performance.[21] In addition, split-complex based algorithms do not have a generic form of their real-valued counterparts, and hence their signal flow-graphs are fundamentally different [220]. In the classification context, early results on Boolean threshold functions and the notion of multiple-valued threshold function can be found in [7, 8].

The problems associated with the choice of complex nonlinearities suitable for nonlinear adaptive filtering in $\mathbb{C}$ have been addressed by Kim and Adali in 2003 [152]. They have identified a class of 'fully complex' activation functions (differentiable and bounded almost everywhere in $\mathbb{C}$ such as tanh), as a suitable choice, and have derived the fully complex back-propagation algorithm [150, 151], which is a generic extension of its real-valued counterpart. They also provide an insight into the character of singularities of fully complex nonlinearities, together with their universal function approximation properties. Uncini *et al.* have introduced a 2D splitting complex activation function [298], and have also applied complex neural networks in the context of blind equalisation [278] and complex blind source separation [259].

Learning algorithms. The first adaptive signal processing algorithm operating completely in $\mathbb{C}$ was the complex least mean square (CLMS), introduced in 1975 by Widrow, Mc Cool and Ball [307] as a natural extension of the real LMS. Work on complex nonlinear architectures, such as complex neural networks (NNs) started much later. Whereas the extension from real LMS to CLMS was fairly straightforward, the extensions of algorithms for nonlinear adaptive filtering from $\mathbb{R}$ into $\mathbb{C}$ have not been trivial. This is largely due to problems associated with the

[20] This is the famous 13th problem of Hilbert, which has been the basis for the development of adaptive models for universal function approximation [56, 125, 126, 155].

[21] Split-complex algorithms cannot calculate the true gradient unless the real and imaginary weight updates are mutually independent. This proves useful, e.g. in communications applications where the data symbols are made orthogonal by design.