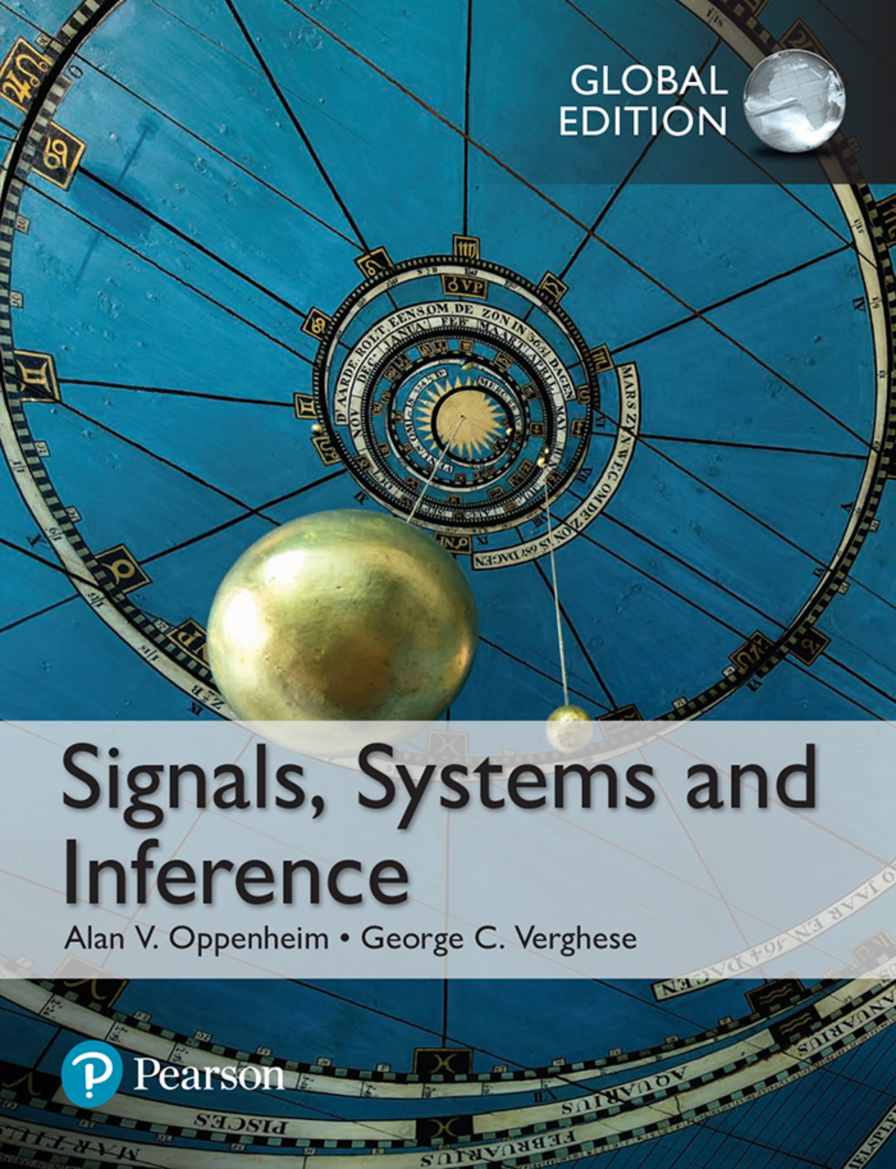
GLOBAL EDITION



Signals, Systems and Inference

Alan V. Oppenheim • George C. Verghese

Pearson

Signals, Systems & Inference

Global Edition

Alan V. Oppenheim & George C. Verghese
Massachusetts Institute of Technology

Pearson
Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Vice President and Editorial Director, ECS: *Marcia J. Horton*
Acquisitions Editor: *Julie Bai*
Executive Marketing Manager: *Tim Galligan*
Marketing Assistant: *Jon Bryant*
Senior Managing Editor: *Scott Disanno*
Program Manager: *Joanne Manning*
Project Editor, Global Edition: *Radhika Raheja*
Acquisitions Editor, Global Edition: *Abhijit Baroi*
Global HE Director of Vendor Sourcing and Procurement: *Diane Hynes*
Director of Operations: *Nick Sklitsis*
Operations Specialist: *Maura Zaldivar-Garcia*
Cover Designer: *Lumina Datamatics*
Manager, Rights and Permissions: *Rachel Youdelman*
Associate Project Manager, Rights and Permissions: *Timothy Nicholls*
Senior Manufacturing Controller, Global Edition: *Kay Holman*
Media Production Manager, Global Edition: *Vikram Kumar*
Full Service Project Management: *Pavithra Jayapaul, Jouve India*

Pearson Education Limited

Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

British Library Cataloguing-in-Publication Data

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

10 9 8 7 6 5 4 3 2 1

ISBN 10: 1292156201
ISBN 13: 9781292156200

Typeset by Jouve India
Printed and bound in Malaysia

We dedicate this book to

Amar Bose, Bernard Gold, and Thomas Stockham

&

George Sr. and Mary Verghese, and Thomas Kailath

These extraordinary people have had a profound impact on our lives and our careers

This page left intentionally blank

Contents

Preface

This book has grown out of an undergraduate course developed and taught by us in MIT's Department of Electrical Engineering and Computer Science. Our course is typically taken by third- and fourth-year undergraduate students from many engineering branches, as well as undergraduate and graduate students from applied science. There are two formal prerequisites for the course, and for this book: an introductory subject in time- and frequency-domain analysis of signals and systems, and an introductory subject in probability. These two subjects are typically taken by most engineering students early in their degree programs. The signals and systems subject almost invariably builds on an earlier course in differential equations, ideally with some basic linear algebra folded into it.

In many engineering departments, students with a strong interest in applied mathematics have then traditionally gone on to a more specialized undergraduate subject in control, signal processing, or communication. In addition to being specialized, such subjects often focus on deterministic signals and systems. Our aim instead was to build broadly on the prerequisite material, folding together signals, systems, and probability in ways that could make our course relevant and interesting to a wider range of students. The course could then serve both as a terminal undergraduate subject and as a sufficiently rigorous basis for more advanced undergraduate subjects or introductory graduate subjects in many engineering and applied science departments.

The course that gave rise to this book teaches students about signals and signal descriptions that are typically new to them, for example, random signals and their characterization through correlation functions and power spectral densities. It introduces them to new kinds of systems and system properties, such as state-space models, reachability and observability, optimum filters, and group delay. And it highlights model-based approaches to inference, particularly in the context of state estimation, signal estimation, and signal detection.

Although some parts of our course are well covered by existing textbooks, we did not find one that fit our needs across the range of topics. This led to lecture notes, which was the easier part, and then eventually this book. In the process, we continually experimented with and refined the content and order of presentation. Along the way we also at times included other material or excluded some that is now back in the book. Among the conclusions of these experiments was that we did not have time in a one-semester class to fold in even basic notions of information theory, despite its central importance to communication systems and, more generally, to inference.

As suggested in the Prologue to this book, signals, systems and probability have been and will continue to be usefully combined in studying fields such as signal processing, control, communication, financial engineering, biomedicine, and many others that involve dynamically varying processes operating in continuous or discrete time, and affected by disturbances, noise, or uncertainty. This premise forms the basis for the overall organization and content of our course and this text.

The book can be thought of as comprising four parts, outlined below. A more detailed overview of the individual chapters is captured in the table of contents. Chapters 1 and 2 present a brief review of the assumed prerequisites in signals and linear time-invariant (LTI) systems, though some portions of the material may be less familiar. A key intent in these chapters is to establish uniform notation and concepts on which to build in the chapters that follow. Chapter 3 discusses the application of some of this prerequisite material in the setting of digital communication by pulse amplitude modulation.

Chapters 4–6 are devoted to state-space models, concentrating on the single-input single-output LTI case. The development is largely built around the eigenmodes of such systems, under the simplifying assumption of distinct natural frequencies. This part of the book introduces the idea of model-based inference in the context of state observers for LTI systems, and examines associated feedback control strategies.

Chapters 7–9 provide a brief review of the assumed probability prerequisites, including estimation and hypothesis testing for static random variables. As with Chapters 1 and 2, we felt it important to set out our notation and perspectives on the concepts while making contact with what students might have encountered in their earlier probability subject. Again, some parts of this material, particularly on hypothesis testing, may be previously unfamiliar to some students.

In Chapters 10–13, we characterize wide-sense stationary random signals, and the outputs that result from LTI filtering of such signals. The associated properties and interpretations of correlation functions and power spectral densities are then used to study canonical signal estimation and signal detection problems. The focus in Chapter 12 is on linear minimum mean square error signal estimation, i.e., Wiener filtering. In Chapter 13, the emphasis is on signal detection for which optimum solutions involve matched filtering.

As is often said, the purpose of a course is to uncover rather than to cover a subject. In this spirit, each chapter includes a final section with some suggestions for further reading. Our intent in these brief sections is not to be exhaustive but rather to suggest the wealth of learning opened up by the material in this text. We have pointed exclusively to books rather than to papers in the research literature, and have in each case listed only a fraction of the books that could have been listed.

Each chapter contains a rich set of problems, which have been divided into Basic, Advanced, and Extension. Basic problems are likely to be easy for most students, while the Advanced problems may be more demanding. The Extension problems often involve material somewhat beyond what is developed in the chapter. Certain problems require simulation or computation using some appropriate computational package. Given the variety and ubiquity of such packages, we have intentionally not attempted to structure the computational exercises around any specific platform.

There is more material in this book than can be taught comfortably in a one-semester course. This allows the instructor or self-learner to choose different routes through the text, and over the years we have experimented with various paths. For a course that is more oriented towards communication or signal processing, Chapters 4, 5 and 6 (state-space models) can be omitted, or addressed only briefly. For a course with more of a control orientation, Chapter 3 (pulse amplitude modulation), Chapter 9 (hypothesis testing) and Chapter 13 (signal detection) can perhaps be considered optional.

A third version of the course, and the one that we currently teach, is outlined in a little more detail below. This version involves two weekly lectures over a semester of approximately thirteen weeks. The lectures are interleaved with an equal number of small-group recitation sections, devoted to more interactive discussion of specific problems that illustrate the lectures and help address the weekly homework. In addition, we staff optional small-group tutorials. Finally an optional evening "common room" that we run several times each week allows students in the class to congregate and interact with each other and with a member of the teaching staff while they work on their homework.

In our teaching in general, we like to emphasize that the homework is intended to provide an occasion for learning and engaging with the concepts and mechanics, rather than being an exam. We recommend that the end-of-chapter problems in this book be approached in the same spirit. In particular, we encourage students to work constructively together, sharing insights and approaches. Our grading of the problems is primarily for feedback to the students and to provide some accountability and motivation. The course does typically have a midterm quiz and a final exam, and many of the end-of-chapter problems in this text were first created as quiz or exam problems. There are also many possibilities for term projects that can grow out of the material in the class, if desired.

An introductory lecture in the same spirit as the Prologue to this text is followed by a brief review of the signals and systems material in Chapter 1. The focus in class is on what might be less familiar from the prerequisite subject, and students are tasked with reviewing the rest on their own, guided by appropriate homework problems. We then move directly to the state-space material in Chapters 4, 5 and 6. Even if students have had some prior exposure to state-space models, there is much that is likely to be new to them here, though they generally relate easily to the material. We have not held students responsible for the more detailed proofs, such as those on eigenvalue placement for LTI observers or state feedback, but do expect them to develop an understanding of the relevant results and how to apply them to small examples. An important lesson from the state-space observer framework is the role of a system model in going from measured signals to inferences about the system.

Our course then turns to probabilistic models and random signals. The probability review in Chapter 7 is mostly woven into lectures covering minimum mean square error (MMSE) and linear MMSE (LMMSE) estimation, which are dealt with in Chapter 8. In order to move more quickly to random signals rather than linger on review of material from the prerequisite probability course, we defer the study of hypothesis testing in Chapter 9 to the end of the course, using it as a lead-in to the signal detection material in Chapter 13. Part of the rationale is also that Chapters 9 and 13 are devoted to making inferences about discrete random quantities, namely the hypotheses, whereas Chapters 8 and 12 on (L)MMSE estimation deal with inferences about continuous random variables. We therefore move directly from Chapter 8 to Chapter 10, studying random signals, i.e., stochastic processes, focusing on the time-domain analysis of wide-sense stationary (WSS) processes, and LTI filtering of such processes.

The topic of power spectral density in Chapter 11 connects back to the development of transforms and energy spectral density in Chapter 1, and also provides the opportunity to refer to relevant sections of Chapter 2 on all-pass filters and spectral factorization. These topics are again important in Chapter 12, on LMMSE (or Wiener) filtering for WSS processes. In most offerings of the course, we omit the full causal Wiener filter development, instead only treating the case of prediction of future values of a process from past values of the same process.

The last part of the course refers strongly back to Chapter 3, using the context of digital communication via pulse amplitude modulation to motivate the hypothesis testing problem. The return to Chapter 3 can also involve reference to the material in Chapter 2 on channel distortions and group delay. The hypothesis testing paradigm is then treated as in Chapter 9. This serves as the foundation for the study of signal detection in the last chapter, Chapter 13.

The breadth of this book, and the different backgrounds we brought to the project, meant that we had much to learn from each other. We also learn each term from the very engaged students, teaching assistants and faculty colleagues who are involved in the course, as well as from the literature on the subjects treated here. This book will have amply met its objectives if it sparks and supports a similar voyage of discovery in its readers, as they construct their own individual re-synthesis of the themes of signals, systems and inference.

Alan V. Oppenheim & George C. Verghese
Cambridge, Massachusetts

The Cover

The choice of images for the front and back covers of both the North American Edition and this Global Edition originated in our desire to suggest some of the book's themes in a visually pleasing and striking way. Our explorations began with images of sundials, clocks, and astrolabes. The astrolabe (www.astrolabes.org), invented over two thousand years ago and used well into the 17th century, was an important instrument for astronomy and navigation. Our search for the front cover of this Global Edition eventually led to the photograph by Frans Lemmens (www.franslemmens.com), taken inside the Eisinga Planetarium (www.planetarium-friesland.nl/en) in Franeker, Holland. This exquisite scale model of the solar system was meticulously built by the amateur astronomer Eise Eisinga in the ceiling of his living room, during the period 1774–1781, and is considered the oldest functioning planetarium.

The image of the dwarf planet Ceres on the back cover of this edition is derived from photographs taken by NASA's spacecraft Dawn (www.dawn.jpl.nasa.gov), which entered into orbit around Ceres in March 2015, after an eight-year journey from our planet. The mastery of signals, systems and inference that humankind has attained in the four centuries since the astrolabe faded from use is represented here: in the precisely controlled launch and trajectory of the Dawn spacecraft – which first included a rendezvous with the asteroid Vesta before moving on to Ceres – and in the subsequent recording, retrieval, and processing of data from it to yield such revealing and awe-inspiring images. But the image also evokes the boundless opportunities for new advances and horizons.

This page left intentionally blank

Acknowledgments

This text has its origins in an MIT subject that was first planned, designed and taught by us over twenty years ago. It has subsequently evolved to its current form through continual experimentation and with many variations of the material and presentation. The subject was conceived as part of the curriculum for a five-year Master of Engineering degree program that was being launched at that time in our Department of Electrical Engineering and Computer Science (EECS). We are grateful to Paul Penfield (as then department head), Jeffrey Shapiro (as associate head) and William Siebert for their part in defining the curricular structure that provided the opening for such a subject. Jeff Shapiro also worked with us on the initial definition of the content. Continued support of the curriculum, and of revisions to it, from subsequent department heads – John Guttag, Rafael Reif, Eric Grimson, Anantha Chandrakasan – and their administrations has been important, and we thank them for their support. More generally, we consider ourselves very fortunate to have had our academic careers develop in this highly collegial and vibrant department. MIT's culture of dedication to teaching and learning informed by research, and the Institute's recognition and celebration of excellence in teaching, have had a significant influence on us.

The staffing of the course, as taught in our department, includes a faculty member who gives two weekly lectures and has overall responsibility for running the course, as well as recitation instructors and teaching assistants who meet regularly with smaller groups of students. Numerous faculty colleagues in our department have collaborated with us over the years, as recitation instructors or as lecturers for the subject. Many students have served as able and enthusiastic teaching assistants. We have also benefited from the help of excellent administrative assistants. We take the opportunity in what follows to thank all these people for their multifaceted contributions to the development and running of the course, to the student experience in the course, and to this text.

In addition to each of us individually and jointly lecturing and overseeing the administration of the course many times, other colleagues who have served in that role are Bernard Lesieutre, Charles Rohrs, Jeffrey Shapiro, Gregory Wornell, and John Wyatt. In the process they have provided valuable feedback on the course content and course notes, as well as bringing new insights and developing new exam and homework problems.

Over the years, we have been privileged to work with a superbly talented and committed roster of faculty and senior graduate students serving as recitation instructors. The recitation instructors who have participated in the teaching of the subject are Jinane Abounadi, Elfar Adalsteinsson, Babak Ayazifar, Duane Boning, Petros Boufounos, John Buck, Mujdat Cetin, Jorge Goncalves, Julie Greenberg, Christoforos Hadjicostis, Peter Hagelstein, Thomas Heldt, Steven Isabelle, Franz Kaertner, James Kirtley, Amos Lapidoth, Bernard Lesieutre, Steve Massaquoi, Shay Maymon, Alexandre Megretski, Jose Moura, Asuman Ozdaglar, Michael Perrott, Rajeev Ram, Charles Rohrs, Melanie Rudoy, Jeffrey Shapiro, Ali Shoeb, William Siebert, Vladimir Stojanovic, Collin Stultz, Russell Tedrake, Mitchell Trott, Thomas Weiss, Alan Willsky, Gregory Wornell, John Wyatt, Laura Zager, and Lizhong Zheng. These colleagues have helped provide a rich experience for the students, and have made many contributions to the content of the course and this text.

Both we and the students in the class have been the beneficiaries of the dedication and energy of the stellar teaching assistants during this period: Irina Abarinov, Abubakar Abid, Anthony Accardi, Chalee Asavathiratham, Thomas Baran, Leighton Barnes, Soosan Beheshti, Ballard Blair, Petros Boufounos, Venkat Chandrasekaran, Jon Chu, Aaron Cohen, Roshni Cooper, Ujjaval Desai, Vijay Divi, Shihab Elborai, Baris Erkmen, Siddhartan Govindasamy, Hanhong Gao, James Geraci, Michael Girone, Carlos Gomez-Uribe, Christoforos Hadjicostis, Andrew Halberstadt, Nicholas Hardy, Everest Huang, Irena Hwang, Zahi Karam, Asif Khan, Alaa Kharbouch, Ashish Khisti, Lohith Kini, Alison Laferriere, Ryan Lang, Danial Lashkari, Adrian Lee, Karen Lee, Durodami Lisk, Karen Livescu, Lorenzo Lorilla, Zhipeng Li, Peter Mayer, Rebecca Mieloszyk, Jose Oscar Mur Miranda, Kirimania Murithi, Akshay Naheta, Kenny Ng, Tri Ngo, Paul Njoroge, Ehimwenma Nosakhare, Uzoma Orji, Tushar Parlikar, Pedro Pinto, Victor Preciado, Andrew Russell, Navid Sabbaghi, Maya Said, Peter Sallaway, Sridevi Sarma, Matthew Secor, Mariam Shanechi, Xiaomeng Shi, Andrew Singer, Lakshminarayan Srinivasan, Brian Stube, Eduardo Sverdlin-Lisker, Kazutaka Takahashi, Sayeed Tasnim, Afsin Ustundag, Kathleen Wage, Tianyu Wang, Keyuan Xu, HoKei Yee, and Laura Zager. Their inputs are reflected in myriad ways throughout this text.

Over the many years of offering this subject, we have been guided by the wisdom of our colleague Frederick Hennie in matters of instructional staffing. Agnes Chow's strategic yet detailed oversight of the EECS department's administrative and financial operations has allowed us and other faculty to focus on our teaching. Lisa Bella, as assistant to the department's education officers, attends almost single-handedly and with incredible responsiveness and good humor to the practical administrative aspects of supporting a hundred professors and over a hundred teaching assistants across the department's teaching enterprise each semester. For administrative assistance with our course in its many offerings, we would like to thank Alecia Batson, Margaret Beucler, Dimonika Bray, Susan Davco, Angela Glass, Vivian Mizuno, Sally Santiago, Darla Secor, Eric Strattman, and Diane Wheeler.

As the class subject has continued to evolve over the two-decade period, the accompanying course notes that ultimately led to this text have also grown and changed. The students in the class have been key participants in that process, through their questions, requests, challenges, suggestions, critiques, and encouragement. It is a continuing privilege to work with the gifted, engaged, thoughtful, and vocal students whom we have in our classrooms at MIT.

We have sometimes said, either ruefully or in jest, that the current text is the fourth edition of a book whose first three editions we never formally published. As any textbook author knows, however, the final phase of producing a polished text from what initially seem to be very good course notes is still a formidable task. Some of our teaching assistants and other students have more recently provided substantial help and feedback in advancing our lecture notes closer to a text. We would like to specifically acknowledge the efforts of Leighton Barnes and Ballard Blair, as well as Manishika Agaskar, Ganesh Ajjanagadde, Michael Mekonnen, Wan-Teh Chang and Guolong Su. For cheerfully, efficiently, and discerningly pulling together and keeping track of all the fragments and versions and edits as we advanced towards a text, we are enormously indebted to Laura von Bosau.

Our department leadership has consistently encouraged us to take the course notes beyond their role as a supplement to the classroom subject and into a published book, so that the material would be more widely and independently accessible. Anantha Chandrakasan's urging in recent years was a key catalyst in making this text happen. Also significant, and greatly appreciated, was the interest from several publishers. Tom Robbins saw the potential early on, and regularly offered helpful advice through the first decade of the course, during his time at Prentice Hall. Phil Meyler generously arranged for detailed feedback at a later stage. Our respect for the vision and integrity of vice president and editorial director Marcia Horton and executive editor Andrew Gilfillan at Pearson were major factors in our choice of publisher; their patience, commitment and confidence in the project meant a lot to us. Special thanks are due to the strong and accommodating editorial and production staff, particularly senior managing editor Scott Disanno at Pearson for his personal attention, and senior project manager Pavithra Jayapaul at Jouve for her outstanding and steady marshaling of the production of the North American edition through its countless details. For their post-publication support of the North American edition and their efforts in connection with the production of the present edition, we are grateful to Julie Bai, Joanne Manning, Michelle Bayman, Sandra Rodriguez and Radhika Raheja at Pearson.

In developing the cover design for the North American edition of the book, which has also informed the design for this edition, it was a pleasure to

work closely with Krista Van Guilder, who was manager of media and design in MIT's interdisciplinary Research Laboratory of Electronics (RLE). RLE is the research home for both of us; the creative environment that it provides for research also impacts our teaching, including the development of this text. The forthright leadership of Yoel Fink, and Jeffrey Shapiro before him, and the exemplary competence and friendliness of the RLE headquarters staff, set the tone for RLE.

Getting to a bound book has naturally included weathering various challenges along the way. Not the least of these was reconciling our sometimes differing opinions, instincts, approaches, or styles on many minor and sometimes major issues. It helped that we started as friends, and as respectful colleagues. And the experience of working so closely and extensively together in coauthoring this text has, happily, deepened that respect and friendship.

In concluding, we express some individual and more personal thoughts and acknowledgments.

Al Oppenheim

Much of the DNA in my contributions to this text derives, both literally and metaphorically, from my mother, as an extraordinary mentor and role model for me. It still astonishes me that as one of ten children in a poor immigrant family, whose parents arrived from Eastern Europe through Ellis Island, she managed to make her way through college and then medical school in the late 1920's. And then how, as a single parent, she very successfully raised three children while working full time in public health. An incredible and inspiring woman.

I landed at MIT, somewhat by accident, as a freshman in 1955, and shortly after wrote a letter home indicating that at the end of the first year I likely would leave for somewhere that was more fun. Clearly, before long MIT became fun and gratifying for me, and has been a wonderful place at which to have spent my entire academic life, first as a student and then as a faculty member. A tremendous expression of gratitude is due to MIT and more specifically to all of my teachers and mentors throughout this entire period at MIT. And, as indicated on the dedication page of this book, in particular and in very special ways to three mentors: Amar Bose, Ben Gold, and Tom Stockham, whose support and encouragement had a truly profound impact on me.

One of the most fortunate days of my life was the day I walked into the office of a then young assistant professor, Amar Bose, and subsequently signed on as his first teaching assistant. And he eventually signed on as my doctoral thesis advisor. What I learned from him about teaching, research, and life over the many decades of our relationship affected me in ways too numerous to describe. He set the highest standards in everything that he did, and his accomplishments as a teacher, an inventor, and an entrepreneur are legendary. Tom Stockham was another young assistant professor whom I met during my doctoral program. His excitement about and enthusiasm for my ideas gave me the courage to pursue them. During his years at MIT as a faculty

member and then as research staff at MIT's Lincoln Laboratory, Tom was one of the pioneers of the then unknown field of digital signal processing. Through that and his later research at the University of Utah, Tom became widely acknowledged as the father of digital audio. Tom was an extraordinary teacher, researcher, practical engineer, and friend. I first met Bernard (Ben) Gold during my early days on the MIT faculty while he was a visiting faculty member in EECS. His work on speech compression was the context for his many pioneering contributions to digital signal processing. Ben's brilliance, creativity, and unassuming style were inspirational to me. He was as eager to learn from those around him as they were from him. Amar, Tom and Ben taught me so many things by example, including the importance of passion and extraordinary standards in every pursuit. Their influence on me is woven into the fabric of my life, my career, and this text. I miss them all, and their spirit remains deeply within me.

As any author knows, textbook writing is a long, difficult, but ultimately rewarding process. Throughout my career I've had the opportunity to write and edit a number of books, and in some cases through two or three editions. In that process, I've had the good fortune of collaborating with other wonderful co-authors in addition to George Verghese, specifically Ron Schafer and Alan Willsky. Such major collaborative projects can often strain relationships, but I'm delighted to say that in all cases, strong bonds and friendships have been the result.

I have often been asked whether I enjoy writing. My response typically has been that "writing is difficult and sometimes painful, but I enjoy *having* written." Projects of this magnitude inevitably require tolerance, patience, support and understanding from family and close friends. I've been incredibly fortunate to have had all of that throughout my career from my wife Phyllis, and from our children Justine and Jason, who have always been the source of tremendous joy. And I'm deeply appreciative of Nora Moran for her special friendship and encouragement (and chicken soup) during the completion of this book.

George Verghese

My parents, George Sr. and Mary, grew up in small towns a mere fifteen miles apart in Kerala, India, but first met each other 2500 miles away in Addis Ababa, Ethiopia, where – young, confident, and adventurous – they had traveled in the early 1950's as teachers. Two further continents later, they continue to set a model for me, of lives lived gracefully. I have everything to thank them for, including the brothers they gave me.

Growing up with physics books to chew on at home surely played a part in landing me at the Indian Institute of Technology, Madras, for undergraduate studies. My favorite professors there, V.G.K. Murti (for network theory) and K. Radhakrishna Rao (for electronic circuits), treated their students with respect, and earned it back many times over with the clarity and integrity of their thinking and teaching, and with their friendly approachability. They are probably why becoming a professor began to seem an attractive proposition to me.

I was fortunate to be introduced to linear system theory by Chi-Tsong Chen at the State University of New York at Stony Brook, and still recall the excitement of my first course taught – and so elegantly – by the author of a textbook. A few months later I drove cross-country for a life-changing period at Stanford, to work under Thomas Kailath. It was an exceptional time to be there, particularly for the opportunity to learn from him as he completed his own text on linear systems, but also for the interactions with his other students, an amazing group. He undoubtedly thought forty years ago that he was only signing on to be my doctoral thesis advisor, but fifteen years later found himself a part of my family. I continue to learn from him on other fronts, and am still in awe of his acuity, bandwidth, energy, and generosity.

When I joined the faculty at MIT, I thought I would try it out for two years to see how I liked it. I've stayed for over 35. It has been a privilege to be affiliated with such an extraordinary institution, and with the people – students, faculty, and staff – who make it so. Working with Al Oppenheim has been a highlight.

My friends and extended family have helped me keep my labors on this text in necessary perspective, and I'm grateful to them for that. They will no doubt be relieved, the next time they ask, to hear that I'm not still working on the same book as the last time they checked. Throughout this, my dear wife Ann has been much more patient and understanding than I had any right to expect. And whenever she hit her limits, she hauled us off for a vacation that I invariably discovered I needed as much as she did. I could not have completed this project without her cheerful forbearance. Our daughters Deia and Amaya, now launched on trajectories of their own devising, keep us and each other smiling; they are our greatest blessings.

Prologue

SIGNALS, SYSTEMS AND INFERENCE

Signals, in the sense that we refer to them in this book, have been of interest at least since the time when human societies began to record and analyze numerical data, for example to track climate, commerce, population, disease, and the movements of celestial bodies. We are continually immersed in signals, registering them through our senses, measuring them through instruments, and analyzing, modifying, and interrelating them.

Systems and signals are intimately connected. In many contexts, it is important to understand the behavior of the underlying systems that generate the signals of interest. Furthermore, the challenges of collecting, interpreting, modeling, transforming, and utilizing signals motivate us to design and implement systems for these purposes, and to generate signals to control and manipulate systems.

Inference, as the term is used in this text, refers to combining prior knowledge and available measurements of signals to draw conclusions in the presence of uncertainty. The prior knowledge may take the form of partially specified models for the measured signals. Inference may be associated with the construction and refinement of such models. The implementation of algorithms for inference can also require designing systems to process the measured signals.

The application of concepts and methods involving signals, systems, and inference in combination is pervasive in science, engineering, medicine, and the social sciences. However, the mathematical, algorithmic, and computational underpinnings often evolve to become largely independent of the specific application. It is this common foundational material that is the focus of this text.

A LITTLE HISTORY

An example of the sophistication attained centuries ago in signals, systems and inference is the astrolabe[1], the most popular astronomical instrument of the medieval world, used for navigation and time keeping in addition to charting the positions of celestial objects. Around 150 AD, Ptolemy of Alexandria described in detail the stereographic projection that forms the basis for the astrolabe; the trigonometric framework for this was developed even earlier, by Hipparchus of Rhodes around 180 BC. The instrument itself made its appearance around 400 AD, and was in widespread use well into the 1600s.

The interplay of signals, systems and inference is also nicely illustrated by Carl Friedrich Gauss's celebrated prediction[2] of the location of the asteroid Ceres, almost a full year after it had been lost to view. Ceres, whose image is on the back cover of this book, is now known to be the largest object in the asteroid belt, and – along with Pluto – is classified as a dwarf planet. The astronomer Giuseppe Piazzi in Palermo discovered the object on New Year's Day of 1801, but was only able to track its motion across the sky for a few degrees of arc before it faded six weeks later in the glare of the sun. There was at the time major interest in the possibility of this being a new planet that had been suspected to exist between Mars and Jupiter. The 24-year-old Gauss, using just three of Piazzi's observations, along with strategic combinations and simplifications of equations derived from Kepler's model of the trajectories of celestial objects, and with many days of hand calculation, was able to generate an estimate of the orbit of Ceres. The predictions made by other astronomers, who had typically assumed circular rather than elliptical orbits, failed to yield sightings of the asteroid. However, successful observations using Gauss's specifications were recorded in early December that year, and again on New Year's Eve. As Gauss put it, he had "restored the fugitive to observation." In later refinements of his method to account for all nineteen of Piazzi's observations rather than just three, and to apply to the motions of other celestial objects, Gauss also brought into play the method of least squares, which he had developed several years earlier. Chapter 8 of this text is devoted to the closely related topic of minimum mean square error estimation of random variables, while Chapter 12 extends this to estimation of random signals.

By 1805, and still motivated by the problem of interpolating measurements of asteroid orbits, Gauss had developed an efficient algorithm to compute the coefficients of finite trigonometric series[3]. He unfortunately never published his algorithm, though it was included in his posthumous collected works sixty years later. Variants of this algorithm were then independently rediscovered by others, as the problem of fitting harmonic series arose in diverse settings, for example to represent variations in barometric pressure or underground temperature, to calculate corrections to compasses on ships, or to model X-ray diffraction data from crystals. The most well known of these variants, commonly referred to collectively as the Fast Fourier Transform (FFT), was published by James Cooley and John Tukey[4] in 1965. Coming at a time when programmable electronic digital computers were beginning to enter routine use in science and engineering, the FFT soon found widespread application, and has had a profound impact.

Many of the foundational concepts and analytical tools discussed throughout this text for both deterministic and probabilistic systems, such as those reviewed in Chapters 1 and 7, have their origins in the work of mathematicians and scientists who lived around the time of Gauss, including Pierre-Simon Laplace and Jean-Baptiste Joseph Fourier, though later contributions also feature prominently, of course. Laplace today is most often associated with the transform that bears his name, but his place in probability theory is considerably more significant, for his 1812 treatise on the subject, and as the “discoverer” of the central limit theorem. Other parts of our text derive more directly from advances made in engineering and applied science since 1800.

The invention of the telegraph in the 1830s sparked a revolution[5] in communication, with subsequent major impact on theory and practice related to all of the topics in this book. It also led to advances in other areas such as transportation and weather prediction, in part because messages could now travel faster than horses, trains, and storms. Within a few years the dots and dashes of Morse code were being transmitted over electrical cables extended between and across continents. Telephony followed in the 1870s, wireless telegraphy and AM radio in the early 1900s, FM radio and television in the 1930s, and radar in the 1940s. Today we have satellite communication, wireless internet, and GPS navigation.

All these transformative technologies exploited and enhanced our ability to work with signals, systems and inference, and were significant catalysts for the creative development of electrical engineering in general. They presented the need to effectively generate electrical signals or electromagnetic waves, to characterize transmission media so that these signals could be propagated through them in predictable ways, to design any necessary filtering and amplification at various intermediate stages, and to develop appropriate signal processing circuits and systems for embedding information at the transmitter and extracting the intended information at the receiver. The modern study of signals and systems in engineering degree programs, with circuits as prime examples of systems, began to take root in the 1930s and ’40s. Some of the notions that we describe in Chapter 2 arose primarily in the context of circuits and transmission lines for communication.

Occurring in parallel with advances in communication were developments relevant to the analysis and design of control systems. Among these were analog computation aimed at the simulation of differential equations that modeled various systems of interest. Though the concepts were described over fifty years earlier, the first practical mechanical implementation was the Differential Analyzer of Vannevar Bush and collaborators around 1930. More flexible and powerful electronic versions, namely analog computers using operational amplifiers, were widely used from the 1950s until they were supplanted by digital computers in the 1980's.

The design of self-regulating devices that utilize feedback dates back to at least around 250 BC, with the water clock of Ctesibius of Alexandria. One of the earliest and most important applications of feedback in the industrial age was James Watt's 1788 centrifugal governor for regulating the speed of steam engines, but it was only in 1868 that James Clerk Maxwell[6] showed how to analyze the dynamic stability of such governors. Feedback control began to be routinely incorporated in engineered systems from the beginning of the 20th century. Much of the associated mathematical theory that is in widespread application today – associated with people such as Harold Black, Harry Nyquist, and Hendrik Bode at Bell Labs in the 1920s and '30s – was actually developed in the context of designing stable and robust electronic amplifiers and oscillators for communication and signal processing. Other work on feedback control was motivated by servomechanism design for regulation in industrial manufacturing, chemical processes, power generation, transportation, and similar settings. Aleksandr Lyapunov's work in the 1890s on the stability of linear and nonlinear dynamic systems that were described in state-space form was not widely known till the 1960s, but is now an essential part of systems and control theory. These state-space models and methods, including the study of equilibrium, stability, measurement-driven simulations for state estimation, and feedback control, are treated in Chapters 4, 5, and 6.

Feedback mechanisms also play an essential role in living systems, as was explicitly described in 1865 by the physiologist Claude Bernard. As the mathematical study of communication and control developed in the early 20th century, Norbert Wiener and colleagues in such diverse fields as psychology, physiology, biology and the social sciences recognized the commonality and importance of feedback in these various disciplines. Their interactions in the 1940's eventually led to Wiener's definition and elaboration in 1948 of cybernetics as the study of control and communication in the animal and machine[7].

The treatment of signals, systems and inference in communication, control and signal processing inherently has to address distortion and errors introduced by non-ideal and poorly characterized components. Feedback is often introduced to overcome precisely such difficulties. A related issue, which inserts uncertainty in the behavior of the system, is that of random disturbances. These can corrupt the signal on a communication channel or at the receiver; can affect the performance of a feedback control system; and can affect the reliability of an inferred outcome. By showing how to model random disturbances in probabilistic terms, and characterizing them in the time and frequency domains, mathematical theory has made a significant impact on these applications. The work of Wiener[8] from the 1920's onward helped to set the foundations for engineering applications in these and related areas. A famous report of his on the extrapolation, interpolation and smoothing of time series[9] was a major advance in bringing the notions of Fourier analysis and stochastic processes into the setting of practical problems in signal processing and inference. Chapter 12, building on Chapters 8, 10 and 11, treats a class of filtering problems associated with Wiener's name, and shows how having a model for a random process provides a basis for filtering and prediction.

Claude Shannon went a step further in his revolutionary 1948 papers[10] that essentially gave birth to information theory. He modeled the communication source itself as a discrete random process, and introduced notions of information, entropy, channel capacity and coding that still form the frame of reference for the field. As noted in the Preface, a treatment of information theory is beyond the scope of this text. However, Shannon's work launched the era of digital communication, and the material we study in Chapter 3 on pulse amplitude modulation, including Nyquist's key contributions, is of considerable practical importance in digital communication. The task of signal detection in noise, addressed in Chapters 9 and 13, is also fundamental in this and many other applications.

As indicated at the beginning of this Prologue, another domain of investigation that has a long history and relationship to the material in this text is the study of time series, carried out not only in the natural sciences – astronomy and climatology, for example – and engineering but also in economics and elsewhere in the social sciences. A typical objective in time series analysis is to use measured noisy data to construct causal dynamic models, which can then be used to infer future values of these signals. There is particular interest in detecting and exploiting any trends or periodicities that might exist in the data. The considerations here are similar to those that motivated the work of Wiener and others, and the mathematical tools overlap, though the time-series literature tends to be more application driven and data centered. For example, the notion of a periodogram, which we encounter in Chapter 11, first appears in this literature, as a tool for detecting underlying periodicity in a random process[11].

The emergence over the past half-century of real-time digital computation capabilities has had major impact on the applications of signals, systems and inference, and has also given rise to new theoretical formulations. An important early example of how real-time computation can fundamentally change the approach to a central problem in signal processing and control is the Kalman filter, which generalized Wiener filtering in several respects and greatly extended its application. The seminal state-space formulation[12] introduced by Rudolf Kalman in 1960 for problems of signal filtering involves recursive least squares estimation of the state of a system whose output represents the signal of interest. The filter runs a computational algorithm in parallel with the operation of the system, with the results of the computation also available for incorporation into a feedback control law. The initial use of the Kalman filter was for navigation applications in the space program, but it is now much more widely applied. The treatment of state observers in Chapter 6 of this text makes connections with the Kalman filter, and the relation to the Wiener filter is outlined in Chapter 12.

A GLANCE AHEAD

Among the most striking developments that the transition to the 21st century has brought to signals, systems and inference is vast distributed and networked computational power, including in small, inexpensive, and mobile packages. Advances in computing, communication, control, and signal processing have resulted in connection and action on scales that were imagined by only a few in the 1960s, at the dawn of the Internet, among them J. C. R. Licklider[13]. A transformational event on the path to making this vision a reality today for so much of humanity was Tim Berners-Lee's invention of the World Wide Web in 1989.

The close coupling of continuous- and discrete-time technologies is of growing importance. Digital signals, communication, and computation commonly mediate interactions among analog physical objects – in automotive systems, entertainment, robotics, human-computer interfaces, avionics, smart-grids, medical instrumentation, and elsewhere. It is also increasingly the case that a given engineered device or component is not easily classified as being intended specifically for communication or control or signal processing or something else; these aspects come together in different combinations at different times. The term "cyber-physical system"[14] is sometimes used to describe the combination of a networked interconnection of embedded computers and the distributed physical processes that they jointly monitor and control.

Our continuing exploration of the universe at both the smallest and largest scales relies in many ways on understanding how to work with signals, systems and inference. The invention of the microscope at the end of the 16th century had profound implications for the development of science at the cellular level and smaller. The invention of the telescope a few years later, at the beginning of the 17th century, similarly enlarged our view of the heavens, and had equally revolutionary consequences. The launch of the Hubble telescope in 1990 has led to our current ability to observe the cosmos at distances of hundreds of millions of light-years. The processing of images from the Hubble telescope incorporates sophisticated extensions of the basic concepts in this text. As one illustration, the techniques of deconvolution, an example of which is examined in Chapter 12, have played an important role in processing of Hubble telescope images, and most critically in initially helping to correct the distortions caused by spherical aberrations in the mirror until it was repaired. In 2003 and 2004, the Hubble telescope captured intriguing images of Ceres. And in March 2015, NASA's Dawn spacecraft, after a journey that lasted eight years, entered the orbit of Ceres, obtaining the most detailed and striking pictures yet of this dwarf planet, including the one that is incorporated into the back cover of this book. We imagine Gauss would be pleased.

Our intention in this book is to address foundational material for applications to signals, systems and inference across a broad set of domains in today's world. These applications are deeply embedded in so many of the systems that we see and use in our everyday lives, and yet are virtually invisible to, and taken for granted by, the casual observer or user. Automotive and entertainment systems, for example, are currently among the largest markets for specialized signal processing systems. Without question, this material will remain foundational for many years to come.

Speculations about the future are always subject to surprises. However, it is certain that new implementation platforms will continue to emerge from advances in such disciplines as quantum physics, materials science, photonics, and biology. And new mathematics will also emerge that will impact the study and application of signals, systems and inference. The novel directions that are opened up by these advances will undoubtedly still derive in part from concepts studied in this book, just as so much of what we use today is rooted in very specific ways on contributions from past centuries. The basic principles and concepts central to this text have a rich historical importance and an even richer future.

NOTES

[1] J. E. Morrison, *The Astrolabe*, Janus 2007 (see also Morrison's rich website, astrolabes.org).

[2] D. Teets and K. Whitehead, "The discovery of Ceres: How Gauss became famous," *Mathematics Magazine*, vol. 72, no. 2, pp. 83–93, April 1999.

[3] M. T. Heideman, D. H. Johnson, and C. S. Burrus, "Gauss and the history of the Fast Fourier Transform," *IEEE Acoustics, Speech and Signal Processing Magazine*, pp. 14–21, October 1984.

[4] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, pp. 297–301, 1965.

[5] The launch of the modern information age, including the birth of telegraphy, is richly and vividly described in J. Gleick's *The Information: A History, A Theory, A Flood,* Vintage Books 2012.

[6] The same J. Clerk Maxwell whose equations launched wireless transmission. An informative description and assessment of his work on analyzing the stability of governors is given by O. Mayr, "Maxwell and the origins of cybernetics," *Isis*, vol. 62, no. 4, pp. 424–444, 1971.

[7] N. Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine*, MIT Press 1948; 2nd edition 1961.

[8] An engaging account of Wiener's life and work can be found in F. Conway and J. Siegelman, *Dark Hero of the Information Age: In Search of Norbert Wiener, the Father of Cybernetics*, Basic Books 2005.

[9] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*, MIT Press 1949 (reprinted in 1964).

[10] See C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press 1949 (reprinted in 1998).

[11] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism*, vol. 3, no. 1, pp. 13–41, 1898. The title and venue of this paper are reflective of the sorts of interests that drove early studies in the time series literature.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82 (series D), pp. 35–45, 1960. Although our text does not include direct treatment of the Kalman filter, it does provide the foundation for reading and understanding this paper. The approachable and tutorial fashion in which the paper is written reflects the fact that it introduces an almost entirely new approach to signal filtering.

[13] J. C. R. Licklider, "Man-computer symbiosis," *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, pp. 4–11, 1960.

[14] A lucid description of such systems and the challenges they present is given in the introduction to E. A. Lee and S. A. Seshia's *Introduction to Embedded Systems: A Cyber-Physical Systems Approach*, edition 1.5, LeeSeshia.org, 2014.

1 Signals and Systems

This text assumes a basic background in the representation of linear, time-invariant systems and the associated continuous-time and discrete-time signals, through convolution, Fourier analysis, Laplace transforms, and z-transforms. In this chapter, we briefly summarize and review this assumed background, in part to establish the notation that we will use throughout the text, and also as a convenient reference for the topics in later chapters.

1.1 SIGNALS, SYSTEMS, MODELS, AND PROPERTIES

Throughout this text we will be considering various classes of signals and systems, developing models for them, and studying their properties.

Signals are represented by real- or complex-valued functions of one or more independent variables. They may be one-dimensional, that is, functions of only one independent variable, or multidimensional. The independent variable may be continuous or discrete. For many of the one-dimensional signals, the independent variable is naturally associated with time although it may not correspond to "real time." When the independent variable is continuous, it is enclosed in curved parentheses, and when discrete in square parentheses to denote an integer variable. For example, $x(t)$ would correspond to a continuous-time (CT) signal and $x[n]$ to a discrete-time (DT) signal. The notations $x(\cdot)$ and $x[\cdot]$ will also be used to refer to the entire signal, suppressing the particular variable t or n used to denote time.

In the first six chapters, we focus entirely on deterministic signals. Starting with Chapter 7, we incorporate stochastic signals, that is, signals drawn from an ensemble of signals, any one of which can be the outcome of a given probabilistic process. To distinguish a signal ensemble representing a random process from a deterministic signal, we will typically use uppercase. For example, $X(t)$ would represent a CT random process whereas $x(t)$ would denote a specific signal in the ensemble. Similarly, $X[n]$ would correspond to a DT random process.

Systems are collections of software or hardware elements, components, or subsystems. A system can be viewed as mapping a set of input signals to a set of output or response signals. A more general view (which we don't incorporate in this text) is that a system is an entity imposing constraints on a designated set of signals without distinguishing specific ones as inputs or outputs. Any particular set of signals that satisfies the constraints is termed a behavior of the system.

Models are (usually approximate) mathematical, software, hardware, linguistic, or other representations of the constraints imposed on a designated set of signals by a system. A model is itself a system because it imposes constraints on the set of signals represented in the model, so we often use the words *system* and *model* interchangeably. However, it can sometimes be important to preserve the distinction between something truly physical and our representations of it mathematically or in a computer simulation.

The difference between representation as a mapping or in behavioral form can be illustrated by considering, for example, Ohm's law for a resistor. Expressed as $v(t) = R\,i(t)$, it suggests current $i(t)$ as an input signal and voltage $v(t)$ as the response, whereas expressed as

$$R\,i(t)/v(t) = 1 \tag{1.1}$$

it is more suggestive of a constraint relating these two signals. Similarly, the resistor-capacitor circuit in Figure 1.1 has constraints among the signals $v(t)$, $i_R(t)$, and $v_C(t)$ imposed by Kirchhoff's laws but does not identify which of the variables are input variables and which are output variables. More broadly, a behavioral representation comprises a listing of the constraints that the signals must satisfy. For example, if a particular system imposed a time-shift constraint between two signals without preference as to which would

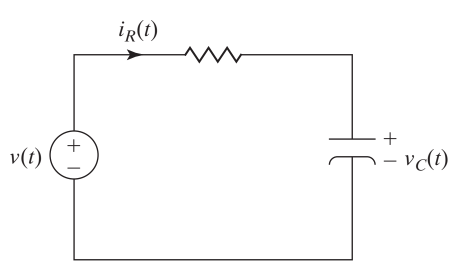


Figure 1.1 Resistor-capacitor circuit.

correspond to the input and which to the output, then a behavioral interpretation would be more appropriate. In this text, we will typically express systems as mappings from inputs to outputs.

The representation of a system or model as a mapping comprises the following: a set of input signals $\{x(\cdot)\}$, each of which can vary within some specified range of possibilities; similarly, a set of output signals $\{y(\cdot)\}$, each of which can vary; and a description of the mapping that uniquely defines the output signals as a function of the input signals.

One way of depicting a system as a mapping is shown in Figure 1.2 for the single-input, single-output CT case, with the interpretation that for each signal in the input set, $T\{\cdot\}$ specifies a mapping to a signal in the output set. Given the input $x(\cdot)$ and the mapping $T\{\cdot\}$, the output $y(\cdot)$ is unique. More commonly, the representation in Figure 1.3 is used to show the input and output signals at some arbitrary time t. With the notation in Figure 1.3, it is important to understand that the mapping $T\{\cdot\}$ is in general a mapping between sets of signals and not a memoryless mapping between a signal value $x(t)$ at a specific time instant to the signal value $y(t)$ at that same time instant. For example, if the system delays the input by t_0, then

$$y(t) = x(t - t_0)\ . \tag{1.2}$$

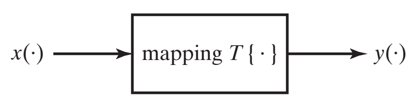


Figure 1.2 Representation of a system as an input-output mapping.

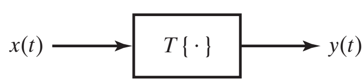


Figure 1.3 Alternative representation of a system as an input-output mapping.

1.1.1 System Properties

For a system specified as a mapping, we use the following definitions of various properties, all of which we assume are familiar. They are stated here for the DT case but are easily modified for the CT case. We also assume a single-input, single-output system in our mathematical representation of the definitions that follow, for notational convenience.

- **Memoryless**: The output at any time instant does not depend on values of the input at any other time instant. The CT delay-by-t_0 system described in Eq. (1.2) is not memoryless. A simple example of a memoryless DT system is one for which

  $$y[n] = x^2[n] \tag{1.3}$$

  for every n.
- **Linear**: The response to an arbitrary linear combination (or "superposition") of input signals is always the same linear combination of the individual responses to these signals.

- **Time-Invariant**: The response to any set of inputs translated arbitrarily in time is always the response to the original set, but translated by the same amount.
- **Linear and Time-Invariant (LTI)**: The system is both linear and time-invariant.
- **Causal**: The output at any instant does not depend on future inputs: for all n_0, $y[n_0]$ does not depend on $x[n]$ for $n > n_0$. Said another way, if $\widehat{x}[n], \widehat{y}[n]$ denotes another input-output pair of the system, with $\widehat{x}[n] = x[n]$ for $n \leq n_0$ where n_0 is fixed but arbitrary, then it must be also true that $\widehat{y}[n] = y[n]$ for $n \leq n_0$.
- **Bounded-Input, Bounded-Output (BIBO) Stable**: The output response to a bounded input is always bounded: $|x[n]| \leq M_x < \infty$ for all n implies that $|y[n]| \leq M_y < \infty$ for all n.

Example 1.1 System Properties

As an example of these system properties, consider the system with input $x[n]$ and output $y[n]$ defined by the relationship

$$y[n] = x[4n+1] \tag{1.4}$$

for all n. We would like to determine whether the system is memoryless, linear, time-invariant, causal, and/or BIBO stable.

Memoryless: A simple counterexample suffices to show that this system is not memoryless. Consider for example $y[n]$ at $n = 0$. From Eq. (1.4), $y[0] = x[1]$ and therefore depends on the value of the input at a time other than at $n = 0$. Consequently it is not memoryless.

Linearity: To check for linearity, we consider two arbitrary input signals, $x_A[n]$ and $x_B[n]$, and compare the output of their linear combination to the linear combination of their individual outputs. From Eq. (1.4), the response $y_A[n]$ to $x_A[n]$ and the response $y_B[n]$ to $x_B[n]$ are respectively (for all n):

$$y_A[n] = x_A[4n+1] \tag{1.5}$$

and

$$y_B[n] = x_B[4n+1]\,. \tag{1.6}$$

If with $x_C[n] = ax_A[n] + bx_B[n]$ for arbitrary a and b the output is $y_C[n] = ay_A[n] + by_B[n]$, then the system is linear. Applying Eq. (1.4) to $x_C[n]$ shows that this holds.

Time Invariance: To check for time invariance, we need to compare the output due to a time-shifted version of $x[n]$ to the time-shifted version of the output due to $x[n]$. The output $y[n]$ resulting from any specific input $x[n]$ is given in Eq. (1.4). The output $\widehat{y}[n]$ results from an input $\widehat{x}[n]$ that is a time-shifted (by n_0) version of the signal $x[n]$. Consequently

$$\widehat{y}[n] = \widehat{x}[4n+1] = x[4n+1+n_0]\,. \tag{1.7}$$

If the system were time-invariant, then $\widehat{y}[n]$ would correspond to shifting $y[n]$ in Eq. (1.4) by n_0, resulting in replacing n by $(n+n_0)$ in Eq. (1.4), which yields

$$y[n+n_0] = x[4n+4n_0+1]\,. \tag{1.8}$$

Since the expressions on the right side of Eqs. (1.7) and (1.8) are not equal, the system is not time-invariant. To illustrate with a specific input, suppose that $x[n]$ is a unit impulse $\delta[n]$, which has the value 1 at $n = 0$ and the value 0 elsewhere. The output $y[n]$ of the system Eq. (1.4) would be $\delta[4n + 1]$, which is zero for all values of n, and $y[n + n_0]$ would likewise always be zero. However, if we consider $x[n + n_0] = \delta[n + n_0]$, the output will be $\delta[4n + 1 + n_0]$, which for $n_0 = 3$ will be 1 at $n = -1$ and zero otherwise.

Causality: Since the output at time $n = 0$ is the input value at $n = 1$, the system is not causal.

BIBO Stability: Since $|y[n]| = |x[4n + 1]|$ and the bound on $|x[n]|$ also bounds $|x[4n + 1]|$, the system is BIBO stable.

1.2 LINEAR, TIME-INVARIANT SYSTEMS

Linear, time-invariant (LTI) systems form the basis for engineering design in many contexts. This class of systems has the advantage of a rich and well-established theory for analysis and design. Furthermore, in many systems that are nonlinear, small deviations from some nominal steady operation are approximately governed by LTI models, so the tools of LTI system analysis and design can be applied incrementally around a nominal operating condition.

1.2.1 Impulse-Response Representation of LTI Systems

A very general way of representing an LTI mapping from an input signal to an output signal is through convolution of the input with the system impulse response. In CT the relationship is

$$y(t) = \int_{-\infty}^{\infty} x(v)h(t - v)\, dv = \int_{-\infty}^{\infty} x(t - \tau)h(\tau)\, d\tau \tag{1.9}$$

where $x(t)$ is the input, $y(t)$ is the output, and $h(t)$ is the unit impulse response of the system. In DT, the corresponding relationship is

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]\, h[n - k] = \sum_{m=-\infty}^{\infty} x[n - m]\, h[m] \tag{1.10}$$

where $h[n]$ is the unit sample (or unit "impulse") response of the system.

The common shorthand notations for the convolution integral in Eq. (1.9) and the convolution sum in Eq. (1.10) are

$$y(t) = x(t) * h(t) \tag{1.11}$$

$$y[n] = x[n] * h[n]\ . \tag{1.12}$$

While these notations can be convenient, they can also easily lead to misinterpretation if not well understood. Alternative notations such as

$$y(t) = (x * h)(t) \tag{1.13}$$

have their advantages and disadvantages. We shall use the notations indicated in Eqs. (1.11) and (1.12) as shorthand for Eqs. (1.9) and (1.10), with the understanding that Eqs. (1.9) and (1.10) are the correct interpretations.

The characterization of LTI systems through convolution is obtained by representing the input signal as a superposition of weighted impulses. In the DT case, suppose we are given an LTI mapping whose impulse response is $h[n]$, that is, when its input is the unit sample or unit "impulse" function $\delta[n]$, its output is $h[n]$. A general input $x[n]$ can be assembled as a sum of scaled and shifted impulses, specifically:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\,\delta[n-k]\,. \tag{1.14}$$

As a consequence of linearity and time invariance, the response $y[n]$ to this input is the sum of the similarly scaled and shifted impulse responses, and is therefore given by Eq. (1.10). What linearity and time invariance have allowed us to do is write the response to a general input in terms of the response to a special input. A similar derivation holds for the CT case.

It may seem that the preceding derivation indicates that all LTI mappings from an input signal to an output signal can be represented through a convolution sum. However, the use of infinite integrals or sums like those in Eqs. (1.9) and (1.10) actually involves some assumptions about the corresponding mapping. We make no attempt here to elaborate on these assumptions. Nevertheless, it is not hard to find "pathological" examples of LTI mappings—not significant for us in this text, or indeed in most engineering models—where the convolution relationship does not hold because these assumptions are violated.

It follows from Eqs. (1.9) and (1.10) that a necessary and sufficient condition for an LTI system to be BIBO stable is that the impulse response be absolutely integrable (CT) or absolutely summable (DT):

$$\text{BIBO stable (CT)} \iff \int_{-\infty}^{\infty} |h(t)|\,dt < \infty \tag{1.15}$$

$$\text{BIBO stable (DT)} \iff \sum_{n=-\infty}^{\infty} |h[n]| < \infty\,. \tag{1.16}$$

It also follows from Eqs. (1.9) and (1.10) that a necessary and sufficient condition for an LTI system to be causal is that the impulse response be zero for $t < 0$ (CT) or for $n < 0$ (DT).

1.2.2 Eigenfunction and Transform Representation of LTI Systems

Exponentials are eigenfunctions of LTI mappings, that is, when the input is an exponential for all time, which we refer to as an "everlasting" exponential, the output is simply a scaled version of the input. Therefore, computing the response to an everlasting exponential reduces to simply multiplying by the appropriate scale factor. Specifically, in the CT case, suppose

$$x(t) = e^{s_0 t} \tag{1.17}$$

for some possibly complex value s_0 (termed the complex frequency). Then from Eq. (1.9)

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)e^{s_0(t-\tau)}d\tau \\ &= H(s_0)e^{s_0 t}\ , \end{aligned} \tag{1.18}$$

where

$$H(s) = \int_{-\infty}^{\infty} h(\tau)e^{-s\tau}\,d\tau\ , \tag{1.19}$$

provided the above integral has a finite value for $s = s_0$ (otherwise the response to the exponential is not well defined). Equation (1.18) demonstrates that $x(t)$ in the form of Eq. (1.17) is an eigenfunction with associated eigenvalue given by $H(s_0)$. Note that Eq. (1.19) is precisely the bilateral Laplace transform of the impulse response, or the transfer function of the system, and the set of values of s in the complex plane for which the above integral takes a finite value constitutes the region of convergence (ROC) of the transform. We discuss the Laplace transform further in Section 1.4.

The fact that the everlasting exponential is an eigenfunction of an LTI system derives directly from the fact that time shifting an everlasting exponential produces the same result as scaling it by a constant factor. In contrast, the one-sided exponential $e^{s_0 t}u(t)$, where $u(t)$ denotes the unit step, is in general not an eigenfunction of an LTI mapping: time shifting a one-sided exponential does not produce the same result as scaling this exponential, as indicated in Example 1.2.

Example 1.2 **Eigenfunctions of LTI Systems**

As demonstrated above, the everlasting complex exponential $e^{j\omega t}$ is an eigenfunction of any LTI system for which the integral in Eq. (1.19) converges at $s = j\omega$, while $e^{j\omega t}u(t)$ is not. Consider, as a simple example, a time delay:

$$y(t) = x(t - t_0)\ . \tag{1.20}$$

The output due to the input $e^{j\omega t}u(t)$ is

$$e^{-j\omega t_0}e^{j\omega t}u(t - t_0)\ .$$

This is not a simple scaling of the input, so $e^{j\omega t}u(t)$ is not in general an eigenfunction of LTI systems.

When $x(t) = e^{j\omega t}$, corresponding to having s_0 take the purely imaginary value $j\omega$ in Eq. (1.17), the input is bounded for all positive and negative time, and the corresponding output is of the form

$$y(t) = H(j\omega)e^{j\omega t} \tag{1.21}$$

provided that $H(s)$ in Eq. (1.19) converges for $s = j\omega$. Here ω is the (real-valued) frequency of the input. From Eq. (1.19), $H(j\omega)$ is given by

$$H(j\omega) = \int_{-\infty}^{\infty} h(t)e^{-j\omega t}\, dt\ . \tag{1.22}$$

The function $H(j\omega)$ in Eq. (1.22) is referred to as the system frequency response, and is also the continuous-time Fourier transform (CTFT) of the impulse response. The integral that defines the CTFT has a finite value for each ω (and can be shown to be a continuous function of ω) if $h(t)$ is absolutely integrable, in other words if

$$\int_{-\infty}^{+\infty} |h(t)|\, dt < \infty\ . \tag{1.23}$$

This condition ensures that $s = j\omega$ is in the ROC of $H(s)$. Comparing Eq. (1.23) and Eq. (1.15), we note that this condition is equivalent to the system being BIBO stable. The CTFT can also be defined for certain classes of signals that are not absolutely integrable, as for $h(t) = (\sin t)/t$ whose CTFT is a rectangle in the frequency domain, but we defer examination of conditions for existence of the CTFT to Section 1.3.

Knowing the response to $e^{j\omega t}$ allows us to also determine the response to a general (real) sinusoidal input of the form

$$x(t) = A\cos(\omega t + \theta) = \frac{A}{2}\left[e^{j(\omega t+\theta)} + e^{-j(\omega t+\theta)}\right]\ . \tag{1.24}$$

Invoking superposition, and assuming $h(t)$ is real so $H(j\omega)$ is conjugate symmetric, some algebra shows that the corresponding output is

$$y(t) = \big|H(j\omega)\big|\, A\cos(\omega t + \theta + \angle H(j\omega))\ . \tag{1.25}$$

Thus the output is again a sinusoid at the same frequency, but scaled in magnitude by the magnitude of the frequency response at the input frequency, and shifted in phase by the angle of the frequency response at the input frequency.

We can similarly examine the eigenfunction property in the DT case. A DT everlasting exponential is a geometric sequence or signal of the form

$$x[n] = z_0^n \tag{1.26}$$

for some possibly complex value z_0, termed the complex frequency. With this DT exponential input, the output of a convolution mapping follows by a simple computation that is analogous to what we showed above for the CT case. Specifically,

$$y[n] = h[n] * x[n] = H(z_0)z_0^n\ , \tag{1.27}$$

where

$$H(z) = \sum_{k=-\infty}^{\infty} h[k]z^{-k}\ , \tag{1.28}$$

provided the above sum has a finite value when $z = z_0$. Note that this sum is precisely the bilateral z-transform of the impulse response, and the set of values of z in the complex plane for which the sum takes a finite value constitutes the ROC of the z-transform. As in the CT case, the one-sided exponential $z_0^n u[n]$ is not in general an eigenfunction. We discuss the z-transform further in Section 1.4.

Again, an important case is when $x[n] = (e^{j\Omega})^n = e^{j\Omega n}$, corresponding to z_0 in Eq. (1.26) having unit magnitude and taking the value $e^{j\Omega}$, where Ω—the (real) "frequency"—denotes the angular position (in radians) around the unit circle in the z-plane. Such an $x[n]$ is bounded for all positive and negative time. Although we use a different symbol, Ω, for frequency in the DT case, to distinguish it from the frequency ω in the CT case, it is not unusual in the literature to find ω used in both CT and DT cases for notational convenience. The corresponding output is

$$y[n] = H(e^{j\Omega})e^{j\Omega n} \tag{1.29}$$

provided that $e^{j\Omega}$ is in the ROC of $H(z)$. From Eq. (1.28), $H(e^{j\Omega})$ is given by

$$H(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} h[n]e^{-j\Omega n} \ . \tag{1.30}$$

The function $H(e^{j\Omega})$ in Eq. (1.30) is the frequency response of the DT system, and is also the discrete-time Fourier transform (DTFT) of the impulse response. The sum that defines the DTFT has a finite value (and can be shown to be a continuous function of Ω) if $h[n]$ is absolutely summable, in other words provided

$$\sum_{n=-\infty}^{\infty} |\, h[n]\, | < \infty \ . \tag{1.31}$$

This condition ensures that $e^{j\Omega}$ is in the ROC of $H(z)$. As in continuous time, this condition is equivalent to the system being BIBO stable. As with the CTFT, the DTFT can be defined for signals that are not absolutely summable; we will elaborate on this in Section 1.3.

Using Eq. (1.29), assuming $h[n]$ is real, and proceeding as in the CT case, it follows that the response to the sinusoidal input

$$x[n] = A\cos(\Omega n + \theta) \tag{1.32}$$

is

$$y[n] = \left|H(e^{j\Omega})\right| A\cos(\Omega n + \theta + \angle H(e^{j\Omega})) \ . \tag{1.33}$$

Note from Eq. (1.30) that the frequency response for DT systems is always periodic, with period 2π. The "low-frequency" response is found in the vicinity of $\Omega = 0$, corresponding to an input signal that is constant for all n. The "high-frequency" response is found in the vicinity of $\Omega = \pm\pi$, corresponding to an input signal $e^{\pm j\pi n} = (-1)^n$ that is the most rapidly varying DT signal possible.

When the input of an LTI system can be expressed as a linear combination of eigenfunctions, for instance (in the CT case)

$$x(t) = \sum_{\ell} a_\ell e^{j\omega_\ell t} , \tag{1.34}$$

then, by linearity, the output is the same linear combination of the responses to the individual exponentials. By the eigenfunction property of exponentials in LTI systems, the response to each exponential involves only scaling by the system's frequency response at the frequency of the exponential. Thus

$$y(t) = \sum_{\ell} a_\ell H(j\omega_\ell) e^{j\omega_\ell t} . \tag{1.35}$$

Similar expressions can be written for the DT case.

1.2.3 Fourier Transforms

A broad class of input signals can be represented as linear combinations of bounded exponentials through the Fourier transform. The synthesis/analysis formulas for the continuous-time Fourier transform (CTFT) are

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\omega)\, e^{j\omega t} d\omega \quad \text{(synthesis)} \tag{1.36}$$

$$X(j\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt \quad \text{(analysis)}. \tag{1.37}$$

Note that Eq. (1.36) expresses $x(t)$ as a linear combination of exponentials, but this weighted combination involves a continuum of exponentials rather than a finite or countable number. If this signal $x(t)$ is the input to an LTI system with frequency response $H(j\omega)$, then by linearity and the eigenfunction property of exponentials the output is the same weighted combination of the responses to these exponentials, that is,

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(j\omega) X(j\omega)\, e^{j\omega t} d\omega . \tag{1.38}$$

By viewing this equation as a CTFT synthesis equation, it follows that the CTFT of $y(t)$ is

$$Y(j\omega) = H(j\omega) X(j\omega) . \tag{1.39}$$

The convolution relationship Eq. (1.9) in the time domain therefore becomes multiplication in the transform domain. Thus, to determine $Y(j\omega)$ at any particular frequency ω_0, we only need to know the Fourier transform of the input at that single frequency, and the frequency response of the system at that frequency. This simple fact serves, in large measure, to explain why the frequency domain is virtually indispensable in the analysis of LTI systems.

The corresponding DTFT synthesis/analysis pair is defined by

$$x[n] = \frac{1}{2\pi} \int_{\langle 2\pi \rangle} X(e^{j\Omega})\, e^{j\Omega n} d\Omega \quad \text{(synthesis)} \tag{1.40}$$

$$X(e^{j\Omega}) = \sum_{n=-\infty}^{\infty} x[n]\, e^{-j\Omega n} \quad \text{(analysis)} \tag{1.41}$$

where the notation $\langle 2\pi \rangle$ on the integral in the synthesis formula denotes integration over any contiguous interval of length 2π. This is because the DTFT is always periodic in Ω with period 2π, a simple consequence of the fact that $e^{j\Omega}$ is periodic with period 2π. Note that Eq. (1.40) expresses $x[n]$ as a weighted combination of a continuum of exponentials.

As in the CT case, it is straightforward to show that if $x[n]$ is the input to an LTI mapping, then the output $y[n]$ has the DTFT

$$Y(e^{j\Omega}) = H(e^{j\Omega})X(e^{j\Omega}) \, . \tag{1.42}$$

1.3 DETERMINISTIC SIGNALS AND THEIR FOURIER TRANSFORMS

In this section, we review the DTFT of deterministic DT signals in more detail and highlight classes of signals that can be guaranteed to have well-defined DTFTs. We shall also devote some attention to the energy density spectrum of signals that have DTFTs. The section will bring out aspects of the DTFT that may not have been emphasized in your earlier signals and systems course. A similar development can be carried out for CTFTs.

1.3.1 Signal Classes and Their Fourier Transforms

The DTFT synthesis and analysis pair in Eqs. (1.40) and (1.41) hold for at least the three large classes of DT signals described below.

Finite-Action Signals

Finite-action signals, which are also called absolutely summable signals or ℓ^1 ("ell-one") signals, are defined by the condition

$$\sum_{k=-\infty}^{\infty} \Big| x[k] \Big| < \infty \, . \tag{1.43}$$

The sum on the left is often called the action of the signal. For these signals, the infinite sum that defines the DTFT is well behaved and the DTFT can be shown to be a continuous function for all Ω. In particular, the values at $\Omega = +\pi$ and $\Omega = -\pi$ are well defined and equal to each other, which need not be the case when signals are not ℓ^1.

Finite-Energy Signals

Finite-energy signals, which are also referred to as square summable or ℓ^2 ("ell-two") signals, are defined by the condition

$$\sum_{k=-\infty}^{\infty} \Big| x[k] \Big|^2 < \infty \, . \tag{1.44}$$

The sum on the left is called the energy of the signal.

In discrete time, an absolutely summable (i.e., ℓ^1) signal is always square summable (i.e., ℓ^2). However, the reverse is not true. For example, consider the signal $(\sin \Omega_c n)/\pi n$ for $0 < \Omega_c < \pi$, with the value at $n = 0$ taken to be Ω_c/π, or consider the signal $(1/n)u[n-1]$, both of which are ℓ^2 but not ℓ^1. If $x[n]$ is such a signal, its DTFT $X(e^{j\Omega})$ can be thought of as the limit for $N \to \infty$ of the quantity

$$X_N(e^{j\Omega}) = \sum_{k=-N}^{N} x[k]e^{-j\Omega k} \tag{1.45}$$

and the resulting limit will typically have discontinuities at some values of Ω. For instance, the transform of $(\sin \Omega_c n)/\pi n$ has discontinuities at $\Omega = \pm\Omega_c$.

Signals of Slow Growth

Signals of slow growth are signals whose magnitude grows no faster than polynomially with the time index, for example, $x[n] = n$ for all n. In this case $X_N(e^{j\Omega})$ in Eq. (1.45) does not converge in the usual sense, but the DTFT still exists as a generalized (or singularity) function; for example, if $x[n] = 1$ for all n, then $X(e^{j\Omega}) = 2\pi\delta(\Omega)$ for $|\Omega| \leq \pi$.

Within the class of signals of slow growth, those of most interest to us are bounded (or ℓ^∞) signals defined by

$$\Big|x[k]\Big| \leq M < \infty \tag{1.46}$$

that is, signals whose amplitude has a fixed and finite bound for all time. Bounded everlasting exponentials of the form $e^{j\Omega_0 n}$, for instance, play a key role in Fourier transform theory. Such signals need not have finite energy, but will have finite average power over any time interval, where average power is defined as total energy over total time.

Similar classes of signals are defined in continuous time. Finite-action (or L^1) signals comprise those that are absolutely integrable, that is,

$$\int_{-\infty}^{\infty} \Big|x(t)\Big|\, dt < \infty\ . \tag{1.47}$$

Finite-energy (or L^2) signals comprise those that are square integrable, that is,

$$\int_{-\infty}^{\infty} \Big|x(t)\Big|^2\, dt < \infty\ . \tag{1.48}$$

In continuous time, an absolutely integrable signal (i.e., L^1) may not be square integrable (i.e., L^2), as is the case, for example, with the signal

$$x(t) = \begin{cases} 1/\sqrt{t} & 0 < t \leq 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.49}$$

However, an L^1 signal that is bounded will also be L^2. As in discrete time, a CT signal that is L^2 is not necessarily L^1, as is the case, for example, with the signal

$$x(t) = \frac{\sin \omega_c t}{\pi t}\ . \tag{1.50}$$

In both continuous time and discrete time, there are many important Fourier transform pairs and Fourier transform properties developed and tabulated in basic texts on signals and systems. For convenience, we include here a brief table of DTFT pairs (Table 1.1) and one of CTFT pairs (Table 1.2). Other pairs are easily derived from these by applying various Fourier transform properties. Note that $\delta[\cdot]$ in the left column in Table 1.1 denotes unit sample functions, while $\delta(\cdot)$ in the right column are unit impulses. Also, the DTFTs in Table 1.1 repeat periodically outside the interval $-\pi < \Omega \leq \pi$.

In general, it is important and useful to be fluent in deriving and utilizing the main transform pairs and properties. In the following subsection we discuss Parseval's identity, a transform property that is of particular significance in our later discussion.

There are, of course, other classes of signals that are of interest to us in applications, for instance growing one-sided exponentials. To deal with such

TABLE 1.1 BRIEF TABLE OF DTFT PAIRS

DT Signal	$\longleftrightarrow$	DTFT for $-\pi < \Omega \leq \pi$
$\delta[n]$	$\longleftrightarrow$	1
$\delta[n-n_0]$	$\longleftrightarrow$	$e^{-j\Omega n_0}$
1 (for all n)	$\longleftrightarrow$	$2\pi\delta(\Omega)$
$e^{j\Omega_0 n}$ $(-\pi < \Omega_0 \leq \pi)$	$\longleftrightarrow$	$2\pi\delta(\Omega-\Omega_0)$
$a^n u[n],\ \lvert a\rvert < 1$	$\longleftrightarrow$	$\dfrac{1}{1-ae^{-j\Omega}}$
$u[n]$	$\longleftrightarrow$	$\dfrac{1}{1-e^{-j\Omega}}+\pi\delta(\Omega)$
$\dfrac{\sin\Omega_c n}{\pi n}$	$\longleftrightarrow$	$\begin{cases}1, & -\Omega_c < \Omega < \Omega_c \\ 0, & \text{otherwise}\end{cases}$
$\left.\begin{matrix}1, & -M \leq n \leq M \\ 0, & \text{otherwise}\end{matrix}\right\}$	$\longleftrightarrow$	$\dfrac{\sin[\Omega(2M+1)/2]}{\sin(\Omega/2)}$

TABLE 1.2 BRIEF TABLE OF CTFT PAIRS

CT Signal	$\longleftrightarrow$	CTFT
$\delta(t)$	$\longleftrightarrow$	1
$\delta(t-t_0)$	$\longleftrightarrow$	$e^{-j\omega t_0}$
1 (for all t)	$\longleftrightarrow$	$2\pi\delta(\omega)$
$e^{j\omega_0 t}$	$\longleftrightarrow$	$2\pi\delta(\omega-\omega_0)$
$e^{-at}u(t), \mathcal{R}e\{a\} > 0$	$\longleftrightarrow$	$\dfrac{1}{a+j\omega}$
$u(t)$	$\longleftrightarrow$	$\dfrac{1}{j\omega}+\pi\delta(\omega)$
$\dfrac{\sin\omega_c t}{\pi t}$	$\longleftrightarrow$	$\begin{cases}1, & -\omega_c < \omega < \omega_c \\ 0, & \text{otherwise}\end{cases}$
$\left.\begin{matrix}1, & -M \leq t \leq M \\ 0, & \text{otherwise}\end{matrix}\right\}$	$\longleftrightarrow$	$\dfrac{\sin\omega M}{\omega/2}$

signals, we make use of z-transforms in discrete time and Laplace transforms in continuous time.

1.3.2 Parseval's Identity, Energy Spectral Density, and Deterministic Autocorrelation

An important property of the Fourier transform is Parseval's identity for ℓ^2 signals. For discrete time, this identity takes the general form

$$\sum_{n=-\infty}^{\infty} x[n]y^*[n] = \frac{1}{2\pi}\int_{<2\pi>} X(e^{j\Omega})Y^*(e^{j\Omega})\,d\Omega \tag{1.51}$$

and for continuous time,

$$\int_{-\infty}^{\infty} x(t)y^*(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)Y^*(j\omega)\,d\omega \tag{1.52}$$

where the superscript symbol * denotes the complex conjugate. Specializing to the case where $y[n] = x[n]$ or $y(t) = x(t)$, we obtain

$$\sum_{n=-\infty}^{\infty} |x[n]|^2 = \frac{1}{2\pi}\int_{<2\pi>} |X(e^{j\Omega})|^2\,d\Omega \tag{1.53}$$

$$\int_{-\infty}^{\infty} |x(t)|^2\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |X(j\omega)|^2\,d\omega\,. \tag{1.54}$$

Parseval's identity allows us to evaluate the energy of a signal by integrating the squared magnitude of its transform. What the identity tells us, in effect, is that the energy of a signal equals the energy of its transform (scaled by $1/2\pi$).

The right-hand sides of Eqs. (1.53) and (1.54) integrate the quantities $|X(e^{j\Omega})|^2$ and $|X(j\omega)|^2$. We denote these quantities by $\overline{S}_{xx}(e^{j\Omega})$ and $\overline{S}_{xx}(j\omega)$:

$$\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2 \tag{1.55}$$

or

$$\overline{S}_{xx}(j\omega) = |X(j\omega)|^2\,. \tag{1.56}$$

These are referred to as the energy spectral density (ESD) of the associated signal because they describe how the energy of the signal is distributed over frequency. To justify this interpretation more concretely, for discrete time, consider applying $x[n]$ to the input of an ideal bandpass filter of frequency response $H(e^{j\Omega})$ that has narrow passbands of unit gain and width Δ centered at $\pm\Omega_0$, as indicated in Figure 1.4. The energy of the output signal must then be the energy of $x[n]$ that is contained in the passbands of the filter. To calculate the energy of the output signal, note that this output $y[n]$ has the transform

$$Y(e^{j\Omega}) = H(e^{j\Omega})X(e^{j\Omega})\,. \tag{1.57}$$

Consequently, by Parseval's identity, the output energy is

$$\begin{aligned}\sum_{n=-\infty}^{\infty} |y[n]|^2 &= \frac{1}{2\pi}\int_{\langle 2\pi\rangle} |Y(e^{j\Omega})|^2\,d\Omega \\ &= \frac{1}{2\pi}\int_{\langle 2\pi\rangle} |H(e^{j\Omega})|^2\,|X(e^{j\Omega})|^2\,d\Omega\,. \end{aligned}\tag{1.58}$$

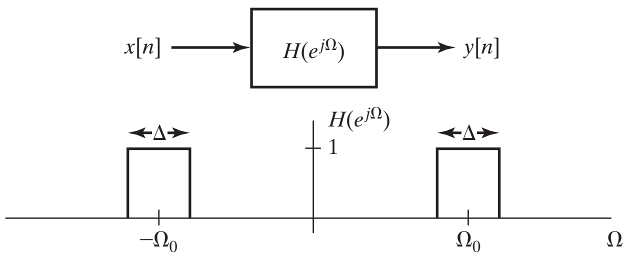


Figure 1.4 Ideal bandpass filter.

Since $|H(e^{j\Omega})|$ is unity in the passband and zero otherwise, Eq. (1.58) reduces to

$$\sum_{n=-\infty}^{\infty} |y[n]|^2 = \frac{1}{2\pi}\int_{\text{passband}} |X(e^{j\Omega})|^2 \, d\Omega$$
$$= \frac{1}{2\pi}\int_{\text{passband}} \overline{S}_{xx}(e^{j\Omega})\, d\Omega \, . \tag{1.59}$$

Thus the energy of $x[n]$ in any frequency band is given by integrating $\overline{S}_{xx}(e^{j\Omega})$ over that band (and scaling by $1/2\pi$). In other words, the energy density of $x[n]$ as a function of Ω is $\overline{S}_{xx}(\Omega)/(2\pi)$ per radian. An exactly analogous discussion can be carried out for CT signals.

Since the ESD $\overline{S}_{xx}(e^{j\Omega})$ is a real function of Ω, an alternate notation for it might be $\mathcal{E}_{xx}(\Omega)$. However, we use the notation $\overline{S}_{xx}(e^{j\Omega})$ in order to make explicit that it is the squared magnitude of $X(e^{j\Omega})$ and also the fact that the ESD for a DT signal is periodic with period 2π.

The ESD also has an important interpretation in the time domain. In discrete time, for example, and assuming $x[n]$ is real, we obtain

$$\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2 = X(e^{j\Omega})X(e^{-j\Omega}) \, . \tag{1.60}$$

Note that $X(e^{-j\Omega})$ is the transform of the time-reversed signal $\overleftarrow{x}[k] = x[-k]$. Thus, since multiplication of transforms in the frequency domain corresponds to convolution of signals in the time domain, we have

$$\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2 \iff x[k] * \overleftarrow{x}[k] = \sum_{n=-\infty}^{\infty} x[n+k]x[n] = \overline{R}_{xx}[k] \, . \tag{1.61}$$

The function $\overline{R}_{xx}[k]$ is referred to as the deterministic autocorrelation function of the signal $x[n]$, and we have just established that the transform of the deterministic autocorrelation function is the energy spectral density $\overline{S}_{xx}(e^{j\Omega})$. A basic Fourier transform property tells us that $\overline{R}_{xx}[0]$, which is the signal energy $\sum_{n=-\infty}^{\infty} x^2[n]$, is the area under the Fourier transform of $\overline{R}_{xx}[k]$, scaled by $1/(2\pi)$, namely the scaled area under $\overline{S}_{xx}(e^{j\Omega}) = |X(e^{j\Omega})|^2$; this, of course, corresponds directly to Eq. (1.53).

The deterministic autocorrelation function measures how alike a signal and its time-shifted version are in a total-squared-error sense. More specifically, in discrete time the total squared error between the signal and its time-shifted version is given by

$$\sum_{n=-\infty}^{\infty} (x[n+k] - x[n])^2 = \sum_{n=-\infty}^{\infty} x^2[n+k] + \sum_{n=-\infty}^{\infty} x^2[n] - 2\sum_{n=-\infty}^{\infty} x[n+k]x[n] = 2(\overline{R}_{xx}[0] - \overline{R}_{xx}[k])\ . \tag{1.62}$$

Since the total squared error is always nonnegative, it follows that $\overline{R}_{xx}[k] \leq \overline{R}_{xx}[0]$, and that the larger the deterministic autocorrelation $\overline{R}_{xx}[k]$ is, the closer the signal $x[n]$ and its time-shifted version $x[n+k]$ are.

Corresponding results hold in continuous time, and in particular

$$\overline{S}_{xx}(j\omega) = |X(j\omega)|^2 \iff x(\tau) * \overleftarrow{x}(\tau) = \int_{-\infty}^{\infty} x(t+\tau)x(t)dt = \overline{R}_{xx}(\tau) \tag{1.63}$$

where $\overline{R}_{xx}(t)$ is the deterministic autocorrelation function of $x(t)$.

1.4 BILATERAL LAPLACE AND *z*-TRANSFORMS

Laplace and z-transforms can be thought of as extensions of Fourier transforms and are useful for a variety of reasons. They permit a transform treatment of certain classes of signals for which the Fourier transform does not converge. They also augment our understanding of Fourier transforms by moving us into the complex plane, where we can apply the theory of complex functions. We begin in Section 1.4.1 with a detailed review of the bilateral z- transform. In Section 1.4.2, we give a short review of the bilateral Laplace transform, paralleling the discussion in Section 1.4.1.

1.4.1 The Bilateral *z*-Transform

The bilateral z-transform is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]z^{-n}\ . \tag{1.64}$$

Here z is a complex variable, which we can also represent in polar form as

$$z = re^{j\Omega}, \quad r \geq 0, \quad -\pi < \Omega \leq \pi \tag{1.65}$$

so

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]r^{-n}e^{-j\Omega n}\ . \tag{1.66}$$

The DTFT corresponds to setting $r = 1$, in which case z takes values on the unit circle. However, there are many useful signals for which the infinite sum does not converge (even in the sense of generalized functions) for z confined to the unit circle. The term z^{-n} in the definition of the z-transform introduces a factor r^{-n} into the infinite sum, which permits the sum to converge (provided r is appropriately restricted) for interesting classes of signals, many of which do not have DTFTs.

More specifically, note from Eq. (1.66) that $X(z)$ can be viewed as the DTFT of $x[n]r^{-n}$. If $r > 1$, then r^{-n} decays geometrically for positive n and grows geometrically for negative n. For $0 < r < 1$, the opposite happens. Consequently, there are many sequences for which $x[n]$ is not absolutely summable, but $x[n]r^{-n}$ is for some range of values of r.

For example, consider $x_1[n] = a^n u[n]$. If $|a| > 1$, this sequence does not have a DTFT. However, for any a, $x[n]r^{-n}$ is absolutely summable provided $r > |a|$. In particular, for example,

$$X_1(z) = 1 + az^{-1} + a^2 z^{-2} + \cdots \tag{1.67}$$

$$= \frac{1}{1 - az^{-1}}, \quad |z| = r > |a| . \tag{1.68}$$

As a second example, consider $x_2[n] = -a^n u[-n-1]$. This signal does not have a DTFT if $|a| < 1$. However, provided $r < |a|$,

$$X_2(z) = -a^{-1}z - a^{-2}z^2 - \cdots \tag{1.69}$$

$$= \frac{-a^{-1}z}{1 - a^{-1}z}, \quad |z| = r < |a| \tag{1.70}$$

$$= \frac{1}{1 - az^{-1}}, \quad |z| = r < |a| . \tag{1.71}$$

The z-transforms of the two distinct signals $x_1[n]$ and $x_2[n]$ above get condensed to the same rational expressions, but for different regions of convergence. Hence the ROC is a critical part of the specification of the transform.

When $x[n]$ is a sum of left-sided and/or right-sided DT exponentials, with each term of the form illustrated in the examples above, then $X(z)$ will be rational in z (or equivalently, in z^{-1}):

$$X(z) = \frac{Q(z)}{P(z)} \tag{1.72}$$

with $Q(z)$ and $P(z)$ being polynomials in z or, equivalently, z^{-1}.

Rational z-transforms are typically depicted by a pole-zero plot in the z-plane, with the ROC appropriately indicated. This information uniquely specifies the signal, apart from a constant amplitude scaling. Note that there can be no poles in the ROC, since the transform is required to be finite in the ROC. z-transforms are often written as ratios of polynomials in z^{-1}. However, the pole-zero plot in the z-plane refers to the roots of the polynomials in z. Also note that if poles or zeros at $z = \infty$ are counted, then any ratio of polynomials always has exactly the same number of poles as zeros.

Region of Convergence

To understand the complex-function properties of the z-transform, we split the infinite sum that defines it into nonnegative-time and negative-time portions. The nonnegative-time or one-sided z-transform is defined by

$$\sum_{n=0}^{\infty} x[n]z^{-n} \tag{1.73}$$

and is a power series in z^{-1}. The convergence of the finite sum $\sum_{n=0}^{N} x[n]z^{-n}$ as $N \to \infty$ is governed by the radius of convergence $R_1 \geq 0$ of the power series. The series converges (absolutely) for each z such that $|z| > R_1$. The resulting function of z is an analytic function in this region, that is, it has a well-defined derivative with respect to the complex variable z at each point in this region, which is what gives the function its nice properties. The series diverges for $|z| < R_1$. The behavior of the sum on the circle $|z| = R_1$ requires closer examination and depends on the particular series; the series may converge (but may not converge absolutely) at all points, some points, or no points on this circle. The region $|z| > R_1$ is referred to as the ROC of the power series for the nonnegative-time part.

Next consider the negative-time part:

$$\sum_{n=-\infty}^{-1} x[n]z^{-n} = \sum_{m=1}^{\infty} x[-m]z^{m} \tag{1.74}$$

which is a power series in z, and has a radius of convergence R_2. The series converges (absolutely) for $|z| < R_2$, which constitutes its ROC; the series is an analytic function in this region. The series diverges for $|z| > R_2$. The behavior on the circle $|z| = R_2$ takes closer examination, and depends on the particular series; and the series may converge (but may not converge absolutely) at all points, some points, or no points on this circle. If $R_1 < R_2$, then the z-transform of $x[n]$ converges (absolutely) for $R_1 < |z| < R_2$; this annular region is its ROC. The transform is analytic in this region. The series that defines the transform diverges for $|z| < R_1$ and $|z| > R_2$. If $R_1 > R_2$, then the z-transform does not exist (for example, for $x[n] = 0.5^n u[-n-1] + 2^n u[n]$). If $R_1 = R_2$, then the transform may exist in a technical sense, but is not useful as a z-transform because it has no ROC. However, if $R_1 = R_2 = 1$, then we may still be able to compute and use a DTFT. For example, for $x[n] = 3$ for all n, or for $x[n] = (\sin \Omega_0 n)/(\pi n)$, the DTFT can be used by incorporating generalized functions such as impulses and step functions in the frequency domain.

Relating the ROC to Signal Properties

For an absolutely summable sequence (such as the impulse response of a BIBO-stable system), that is, an ℓ^1-signal, the unit circle must lie in the ROC or must be a boundary of the ROC. Conversely, we can conclude that a signal is ℓ^1 if the ROC contains the unit circle because the transform converges absolutely in its ROC. If the unit circle constitutes a boundary of the ROC, then further analysis is generally needed to determine if the signal is ℓ^1. Rational transforms always have a pole on the boundary of the ROC, as elaborated on below, so if the unit circle is on the boundary of the ROC of a rational transform, then there is a pole on the unit circle and the signal cannot be ℓ^1.

For a right-sided signal, it is the case that $R_2 = \infty$, that is, the ROC extends everywhere in the complex plane outside the circle of radius R_1, up to (and perhaps including) ∞. The ROC includes ∞ if the signal is zero for negative time.

We can state a converse result if, for example, we know the signal comprises only sums of one-sided exponentials of the form obtained when inverse transforming a rational transform. In this case, if $R_2 = \infty$, then the signal must be right-sided; if the ROC includes ∞, then the signal must be causal, that is, zero for $n < 0$.

For a left-sided signal, $R_1 = 0$, that is, the ROC extends inward from the circle of radius R_2, up to (and perhaps including) zero. The ROC includes $z = 0$ if the signal is zero for positive time.

In the case of signals that are sums of one-sided exponentials, we have the converse: if $R_1 = 0$, then the signal must be left-sided; if the ROC includes $z = 0$, then the signal must be anticausal, that is, zero for $n > 0$.

As indicated earlier, the ROC cannot contain poles of the z-transform because poles are values of z where the transform has infinite magnitude, while the ROC comprises values of z where the transform converges. For signals with *rational* transforms, one can use the fact that such signals are sums of one-sided exponentials to show that the possible boundaries of the ROC are in fact precisely determined by the locations of the poles. Specifically:

(a) The outer bounding circle of the ROC in the rational case contains a pole and/or has radius ∞. If the outer bounding circle is at infinity, then (as we have already noted) the signal is right-sided, and is in fact causal if there is no pole at ∞.

(b) The inner bounding circle of the ROC in the rational case contains a pole and/or has radius 0. If the inner bounding circle reduces to the point 0, then (as we have already noted) the signal is left-sided, and is in fact anticausal if there is no pole at 0.

The Inverse z-Transform

One method for inverting a rational z-transform is using a partial fraction expansion, then either directly recognizing the inverse transform of each term in the partial fraction representation or expanding the term in a power series that converges for z in the specified ROC. For example, a term of the form

$$\frac{1}{1 - az^{-1}} \tag{1.75}$$

can be expanded in a power series in az^{-1} if $|a| < |z|$ for z in the ROC, and expanded in a power series in $a^{-1}z$ if $|a| > |z|$ for z in the ROC. Carrying out this procedure for each term in a partial fraction expansion, we find that the signal $x[n]$ is a sum of left-sided and/or right-sided exponentials. For nonrational transforms, where there may not be a partial fraction expansion to simplify the process, it is still reasonable to attempt the inverse transformation by expansion into a power series consistent with the given ROC.

Although we will generally use partial fraction or power series methods to invert z-transforms, there is an explicit formula that is similar to that of the inverse DTFT, specifically,

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(z)z^n \, d\Omega\Big|_{z=\bar{r}e^{j\Omega}} \tag{1.76}$$

where the constant $\bar{r}$ is chosen to place z in the ROC. This is not the most general inversion formula, but is sufficient for us, and shows that $x[n]$ is expressed as a weighted combination of DT exponentials.

As is the case for Fourier transforms, there are many useful z-transform pairs and properties developed and tabulated in basic texts on signals and systems. Appropriate use of transform pairs and properties is often the basis for obtaining the z-transform or the inverse z-transform of many other signals.

1.4.2 The Bilateral Laplace Transform

As with the z-transform, the Laplace transform is introduced in part to handle important classes of signals that do not have CTFTs, but it also enhances our understanding of the CTFT. The definition of the Laplace transform is

$$X(s) = \int_{-\infty}^{\infty} x(t)\, e^{-st}\, dt \tag{1.77}$$

where s is a complex variable, $s = \sigma + j\omega$. The Laplace transform can thus be thought of as the CTFT of $x(t)\, e^{-\sigma t}$. With σ appropriately chosen, the integral in Eq. (1.77) can exist even for signals that have no CTFT.

The development of the Laplace transform parallels closely that of the z-transform in the preceding section, but with e^{σ} playing the role that r did in Section 1.4.1. The interior of the set of values of s for which the defining integral converges, as the limits on the integral approach $\pm\infty$, comprises the ROC for the transform $X(s)$. The ROC is now determined by the minimum and maximum allowable values of σ, say σ_1 and σ_2 respectively. We refer to σ_1, σ_2 as abscissas of convergence. The corresponding ROC is a vertical strip between σ_1 and σ_2 in the complex plane, $\sigma_1 < \mathrm{Re}\{s\} < \sigma_2$. Equation (1.77) converges absolutely within the ROC; convergence at the left and right bounding vertical lines of the strip has to be separately examined. Furthermore, the transform is analytic (that is, differentiable as a complex function) throughout the ROC. The strip may extend to $\sigma_1 = -\infty$ on the left, and to $\sigma_2 = +\infty$ on the right. If the strip collapses to a line (so that the ROC vanishes), then the Laplace transform is not useful (except if the line happens to be the $j\omega$ axis, in which case a CTFT analysis may perhaps be recovered).

For example, consider $x_1(t) = e^{at}u(t)$; the integral in Eq. (1.77) evaluates to $X_1(s) = 1/(s-a)$ provided $\mathrm{Re}\{s\} > a$. On the other hand, for $x_2(t) = -e^{at}u(-t)$, the integral in Eq. (1.77) evaluates to $X_2(s) = 1/(s-a)$ provided $\mathrm{Re}\{s\} < a$. As with the z-transform, note that the expressions for the transforms above are identical; they are distinguished by their distinct regions of convergence.

The ROC may be associated with properties of the signal. For example, for absolutely integrable signals, also referred to as L^1 signals, the integrand in the definition of the Laplace transform is absolutely integrable on the $j\omega$ axis, so the $j\omega$ axis is in the ROC or on its boundary. In the other direction, if the $j\omega$ axis is strictly in the ROC, then the signal is L^1, because the integral converges absolutely in the ROC. Recall that a system has an L^1 impulse response if and only if the system is BIBO stable, so the result here is relevant to discussions of stability: if the $j\omega$ axis is strictly in the ROC of the system function, then the system is BIBO stable.

For right-sided signals, the ROC is some right half-plane (i.e., all s such that $\mathrm{Re}\{s\} > \sigma_1$). Thus the system function of a causal system will have an ROC that is some right half-plane. For left-sided signals, the ROC is some left half-plane. For signals with rational transforms, the ROC contains no poles, and the boundaries of the ROC will have poles. Since the location of the ROC of a transfer function relative to the imaginary axis relates to BIBO stability, and since the poles identify the boundaries of the ROC, the poles relate to stability. In particular, a system with a right-sided impulse response (e.g., a causal system) will be stable if and only if all its poles are finite and in the left half-plane, because this is precisely the condition that allows the ROC to contain the entire imaginary axis. Also note that a signal with a rational transform and no poles at infinity is causal if and only if it is right-sided.

A further property worth recalling is connected to the fact that exponentials are eigenfunctions of LTI systems. If we denote the Laplace transform of the impulse response $h(t)$ of an LTI system by $H(s)$, then $e^{s_0 t}$ at the input of the system yields $H(s_0)\, e^{s_0 t}$ at the output, provided s_0 is in the ROC of the transfer function.

1.5 DISCRETE-TIME PROCESSING OF CONTINUOUS-TIME SIGNALS

Many modern systems for applications such as communication, entertainment, navigation, and control are a combination of CT and DT subsystems, exploiting the inherent properties and advantages of each. In particular, the DT processing of CT signals is common in such applications, and we describe the essential ideas behind such processing here. As with the earlier sections, we assume that this discussion is primarily a review of familiar material, included here to establish notation and for convenient reference from later chapters in this text. In this section, and throughout this text, we will often relate the CTFT of a CT signal and the DTFT of a DT signal obtained from samples of the CT signal. We will use the subscripts c and d when necessary to help keep clear which signals are CT and which are DT.

1.5.1 Basic Structure for DT Processing of CT Signals

Figure 1.5 depicts the basic structure of this processing, which involves continuous-to-discrete (C/D) conversion to obtain a sequence of samples of the CT input signal; followed by DT filtering to produce a sequence of samples of the desired CT output; then discrete-to-continuous (D/C) conversion to reconstruct this desired CT output signal from the sequence of samples. We will often restrict ourselves to conditions such that the overall system in Figure 1.5 is equivalent to an LTI CT system. The necessary conditions for this typically include restricting the DT filtering to LTI processing by a system with frequency response $H_d(e^{j\Omega})$, and also requiring that the input $x_c(t)$ be appropriately bandlimited. To satisfy the latter requirement, it is typical to precede the structure in the figure by a filter whose purpose is to ensure that $x_c(t)$ is essentially bandlimited. While this filter is often referred to as an anti-aliasing filter, we can often allow some aliasing in the C/D conversion if the DT system removes the aliased components; the overall system can then still be a CT LTI system.

The ideal C/D converter in Figure 1.5 has as its output a sequence of samples of $x_c(t)$ with a specified sampling interval T_1, so that the DT signal is $x_d[n] = x_c(nT_1)$. Conceptually, therefore, the ideal C/D converter is straightforward. A practical analog-to-digital (A/D) converter also quantizes the signal to one of a finite set of output levels. However, in this text we do not consider the additional effects of quantization.

In the frequency domain, the CTFT of $x_c(t)$ and the DTFT of $x_d[n]$ can be shown to be related by

$$X_d(e^{j\Omega})\Big|_{\Omega=\omega T_1} = \frac{1}{T_1}\sum_k X_c\left(j\omega - jk\frac{2\pi}{T_1}\right) . \tag{1.78}$$

When $x_c(t)$ is sufficiently bandlimited so that

$$X_c(j\omega) = 0, \qquad |\omega| \geq \frac{\pi}{T_1} \tag{1.79}$$

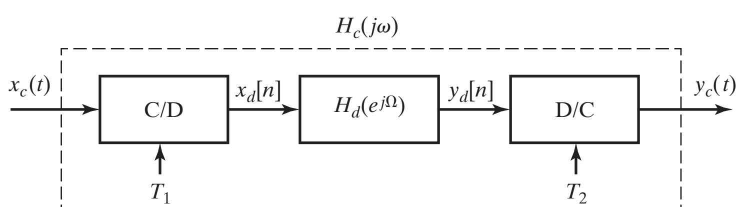


Figure 1.5 DT processing of CT signals.

that is, when the sampling is at or above the Nyquist rate, then Eq. (1.78) can be rewritten as

$$X_d(e^{j\Omega})\bigg|_{\Omega=\omega T_1} = \frac{1}{T_1}X_c(j\omega)\,, \qquad |\omega| < \pi/T_1 \tag{1.80a}$$

or equivalently

$$X_d(e^{j\Omega}) = \frac{1}{T_1}X_c\left(j\frac{\Omega}{T_1}\right)\,, \qquad |\Omega| < \pi\,. \tag{1.80b}$$

Note that $X_d(e^{j\Omega})$ is extended periodically outside the interval $|\Omega| < \pi$.

The ideal D/C converter in Figure 1.5 is defined through the interpolation relation

$$y_c(t) = \sum_n y_d[n]\frac{\sin\left(\pi\ (t - nT_2)\ /T_2\right)}{\pi(t - nT_2)/T_2}\,, \tag{1.81}$$

which shows that $y_c(nT_2) = y_d[n]$. Since each term in the above sum is bandlimited to $|\omega| < \pi/T_2$, the CT signal $y_c(t)$ is also bandlimited to this frequency range, so this D/C converter is more completely referred to as the ideal bandlimited interpolating converter. The C/D converter in Figure 1.5, under the assumption Eq. (1.79), is similarly characterized by the fact that the CT signal $x_c(t)$ is the ideal bandlimited interpolation of the DT sequence $x_d[n]$.

Because $y_c(t)$ is bandlimited and $y_c(nT_2) = y_d[n]$, analogous relations to Eq. (1.80) hold between the DTFT of $y_d[n]$ and the CTFT of $y_c(t)$:

$$Y_d(e^{j\Omega})\bigg|_{\Omega=\omega T_2} = \frac{1}{T_2}Y_c(j\omega)\,, \qquad |\omega| < \pi/T_2 \tag{1.82a}$$

or equivalently

$$Y_d(e^{j\Omega}) = \frac{1}{T_2}Y_c\left(j\frac{\Omega}{T_2}\right)\,, \qquad |\Omega| < \pi\,. \tag{1.82b}$$

Figure 1.6 shows one conceptual representation of the ideal D/C converter. This figure interprets Eq. (1.81) to be the result of evenly spacing a sequence of impulses at intervals of T_2—the reconstruction interval—with impulse strengths given by the $y_d[n]$, then filtering the result by an ideal low-pass filter $L(j\omega)$ with gain T_2 in the passband $|\omega| < \pi/T_2$. This operation

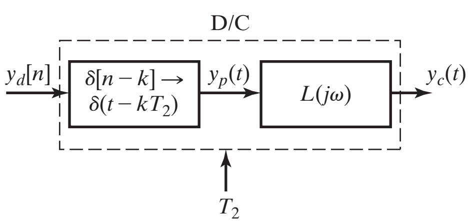


Figure 1.6 Conceptual representation of processes that yield ideal D/C conversion, interpolating a DT sequence into a bandlimited CT signal using reconstruction interval T_2.

produces the bandlimited CT signal $y_c(t)$ that interpolates the specified sequence values $y_d[n]$ at the instants $t = nT_2$, that is, $y_c(nT_2) = y_d[n]$.

1.5.2 DT Filtering and Overall CT Response

We now assume, unless stated otherwise, that $T_1 = T_2 = T$. If in Figure 1.5 the bandlimiting constraint of Eq. (1.79) is satisfied, and if we set $y_d[n] = x_d[n]$, then $y_c(t) = x_c(t)$. More generally, when the DT system in Figure 1.5 is an LTI DT filter with frequency response $H_d(e^{j\Omega})$, so

$$Y_d(e^{j\Omega}) = H_d(e^{j\Omega})X_d(e^{j\Omega}) \,, \tag{1.83}$$

and provided any aliased components of $x_c(t)$ are eliminated by $H_d(e^{j\Omega})$, then assembling Eqs. (1.80), (1.82), and (1.83) yields

$$Y_c(j\omega) = H_d(e^{j\Omega})\Big|_{\Omega=\omega T} X_c(j\omega) \,, \quad |\omega| < \pi/T \,. \tag{1.84}$$

The action of the overall system is thus equivalent to that of a CT filter whose frequency response is

$$H_c(j\omega) = H_d(e^{j\Omega})\Big|_{\Omega=\omega T} \,, \qquad |\omega| < \pi/T \,. \tag{1.85}$$

In other words, under the bandlimiting and sampling rate constraints mentioned above, the overall system behaves as an LTI CT filter, and the response of this filter is related to that of the embedded DT filter through a simple frequency scaling. The sampling rate can be lower than the Nyquist rate, provided that the DT filter eliminates any aliased components.

If we wish to use the system in Figure 1.5 to implement a CT LTI filter with frequency response $H_c(j\omega)$, we choose $H_d\big(e^{j\Omega}\big)$ according to Eq. (1.85), provided that $x_c(t)$ is appropriately bandlimited. If we define $H_c(j\omega) = 0$ for $|\omega| \geq \pi/T$, then Eq. (1.85) also corresponds to the following relation between the DT and CT impulse responses:

$$h_d[n] = T\, h_c(nT) \,. \tag{1.86}$$

The DT filter is therefore a sampled version of the CT filter. When $x_c(t)$ and $H_d(e^{j\Omega})$ are not sufficiently bandlimited to avoid aliased components in $y_d[n]$, then the overall system in Figure 1.5 is no longer time-invariant. It is, however, still linear since it is a cascade of linear subsystems.

The following two examples illustrate the use of Eq. (1.85) as well as Figure 1.5, both for DT processing of CT signals and for interpretation of two important DT systems.

Example 1.3 Digital Differentiator

In this example we wish to implement a CT differentiator using a DT system in the configuration of Figure 1.5. We need to choose $H_d\left(e^{j\Omega}\right)$ so that $y_c(t) = \frac{dx_c(t)}{dt}$, assuming that $x_c(t)$ is bandlimited to π/T. The desired overall CT frequency response is therefore

$$H_c(j\omega) = \frac{Y_c(j\omega)}{X_c(j\omega)} = j\omega \,. \tag{1.87}$$

Consequently, using Eq. (1.85) we choose $H_d(e^{j\Omega})$ such that

$$H_d(e^{j\Omega})\Big|_{\Omega=\omega T} = j\omega \,, \qquad |\omega| < \frac{\pi}{T} \tag{1.88a}$$

or equivalently

$$H_d(e^{j\Omega}) = j\Omega/T \,, \qquad |\Omega| < \pi \,. \tag{1.88b}$$

A DT system with the frequency response in Eq. (1.88b) is commonly referred to as a digital differentiator. To understand the relation between the input $x_d[n]$ and output $y_d[n]$ of the digital differentiator, note that $y_c(t)$—which is the bandlimited interpolation of $y_d[n]$—is the derivative of $x_c(t)$, and $x_c(t)$ in turn is the bandlimited interpolation of $x_d[n]$. It follows that $y_d[n]$ can, in effect, be thought of as the result of sampling the derivative of the bandlimited interpolation of $x_d[n]$.

Example 1.4 Half-Sample Delay

In designing DT systems, a phase factor of the form $e^{-j\alpha\Omega}$, $|\Omega| < \pi$, is often included or required. When α is an integer, this has a straightforward interpretation: it corresponds simply to an integer shift of the time sequence by α. When α is not an integer, the interpretation is not as immediate, since a DT sequence can only be directly shifted by integer amounts.

In this example we consider the case of $\alpha = \frac{1}{2}$, referred to as a half-sample delay. To provide an interpretation, we consider the implications of choosing the DT system in Figure 1.5 to have frequency response

$$H_d(e^{j\Omega}) = e^{-j\Omega/2} \,, \qquad |\Omega| < \pi \,. \tag{1.89}$$

Whether or not $x_d[n]$ explicitly arose by sampling a CT signal, we can associate $x_d[n]$ with its bandlimited interpolation $x_c(t)$ for any specified sampling or reconstruction interval T. Similarly, we can associate $y_d[n]$ with its bandlimited interpolation $y_c(t)$ using the reconstruction interval T. With $H_d\left(e^{j\Omega}\right)$ given by Eq. (1.89), the equivalent CT frequency response relating $y_c(t)$ to $x_c(t)$ is

$$H_c(j\omega) = e^{-j\omega T/2} \tag{1.90}$$

representing a time delay of $T/2$, which is half the sample spacing; consequently, $y_c(t) = x_c(t - T/2)$. We therefore conclude that for a DT system with frequency response given by Eq. (1.89), the DT output $y_d[n]$ corresponds to samples of the half-sample delay of the bandlimited interpolation of the input sequence $x_d[n]$. Note that in this interpretation the choice for the value of T is immaterial.

The preceding interpretation allows us to find the unit sample (or impulse) response of the half-sample delay system through a simple argument. If $x_d[n] = \delta[n]$, then $x_c(t)$ must be the bandlimited interpolation of this (with some T that we could have specified to take any particular value), so

$$x_c(t) = \frac{\sin(\pi t/T)}{\pi t/T} \tag{1.91}$$

and therefore

$$y_c(t) = \frac{\sin\Big(\pi(t-(T/2))/T\Big)}{\pi(t-(T/2))/T} \tag{1.92}$$

which shows that the desired unit sample response is

$$y_d[n] = h_d[n] = \frac{\sin\Big(\pi(n-(1/2))\Big)}{\pi(n-(1/2))}\,. \tag{1.93}$$

This discussion of a half-sample delay also generalizes in a straightforward way to any integer or non-integer choice for the value of α.

1.5.3 Nonideal D/C Converters

In Section 1.5.1 we defined the ideal D/C converter through the bandlimited interpolation formula Eq. (1.81), also illustrated in Figure 1.6, which corresponds to processing a train of impulses with strengths equal to the sequence values $y_d[n]$ through an ideal low-pass filter. A more general class of D/C converters, which includes the ideal converter as a particular case, creates a CT signal $y_c(t)$ from a DT signal $y_d[n]$ according to the following:

$$y_c(t) = \sum_{n=-\infty}^{\infty} y_d[n]\,p(t-nT) \tag{1.94}$$

where $p(t)$ is some selected basic pulse and T is the reconstruction interval or pulse repetition interval. This too can be seen as the result of processing an impulse train of sequence values through a filter, but a filter that has impulse response $p(t)$ rather than that of the ideal low-pass filter. The CT signal $y_c(t)$ is thus constructed by adding together shifted and scaled versions of the basic pulse; the number $y_d[n]$ scales $p(t-nT)$, which is the basic pulse delayed by nT. Note that the ideal bandlimited interpolating converter of Eq. (1.81) is obtained by choosing

$$p(t) = \frac{\sin(\pi t/T)}{\pi t/T}\,. \tag{1.95}$$

In Chapter 3, we will discuss the interpretation of Eq. (1.94) as pulse-amplitude modulation (PAM) for communicating DT information over a CT channel.

The relationship in Eq. (1.94) can also be described quite simply in the frequency domain. Taking the CTFT of both sides, denoting the CTFT of $p(t)$

by $P(j\omega)$, and using the fact that delaying a signal by t_0 in the time domain corresponds to multiplication by $e^{-j\omega t_0}$ in the frequency domain, we get

$$\begin{aligned} Y_c(j\omega) &= \left(\sum_{n=-\infty}^{\infty} y_d[n]\, e^{-jn\omega T}\right) P(j\omega) \\ &= Y_d(e^{j\Omega})\Big|_{\Omega=\omega T} P(j\omega)\,. \end{aligned} \tag{1.96}$$

In the particular case where $p(t)$ is the sinc pulse in Eq. (1.95), with transform $P(j\omega)$ that has the constant value T for $|\omega| < \pi/T$ and 0 outside this band, we recover the relation in Eq. (1.82).

In practice, the ideal frequency characteristic can only be approximated, with the accuracy of the approximation often related to cost of implementation. A commonly used simple approximation is the (centered) zero-order hold (ZOH), specified by the choice

$$p_z(t) = \begin{cases} 1 & \text{for } |t| < (T/2) \\ 0 & \text{elsewhere.} \end{cases} \tag{1.97}$$

This D/C converter holds the value of the DT signal at time n, namely the value $y_d[n]$, for an interval of length T centered at nT in the CT domain, as illustrated in Figure 1.7. The centered ZOH is of course noncausal, but is easily replaced with the noncentered causal ZOH, for which the basic pulse is

$$p_{z'}(t) = \begin{cases} 1 & \text{for } 0 \le t < T \\ 0 & \text{elsewhere.} \end{cases} \tag{1.98}$$

Such ZOH converters are commonly used.

Another common choice is a centered first-order hold (FOH), for which the basic pulse $p_f(t)$ is triangular as shown in Figure 1.8. Use of the FOH represents linear interpolation between the sequence values. Of course, the use of the ZOH and FOH will not be equivalent to exact bandlimited interpolation as required by the Nyquist sampling theorem. The transform of the centered ZOH pulse is

$$P_z(j\omega) = T\frac{\sin(\omega T/2)}{\omega T/2} \tag{1.99}$$

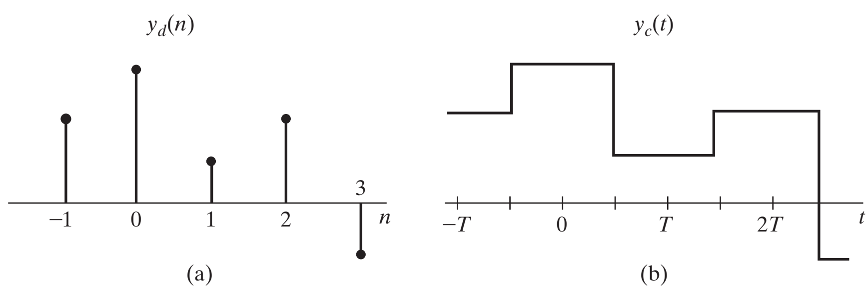


Figure 1.7 A centered zero-order hold (ZOH): (a) DT sequence; (b) the result of applying the centered ZOH to (a).

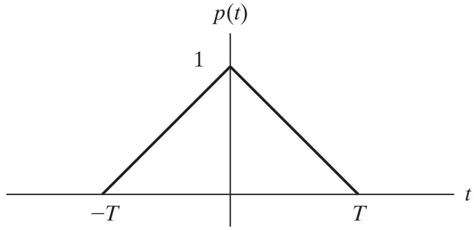


Figure 1.8 Basic pulse $p_f(t)$ for centered first-order hold (FOH).

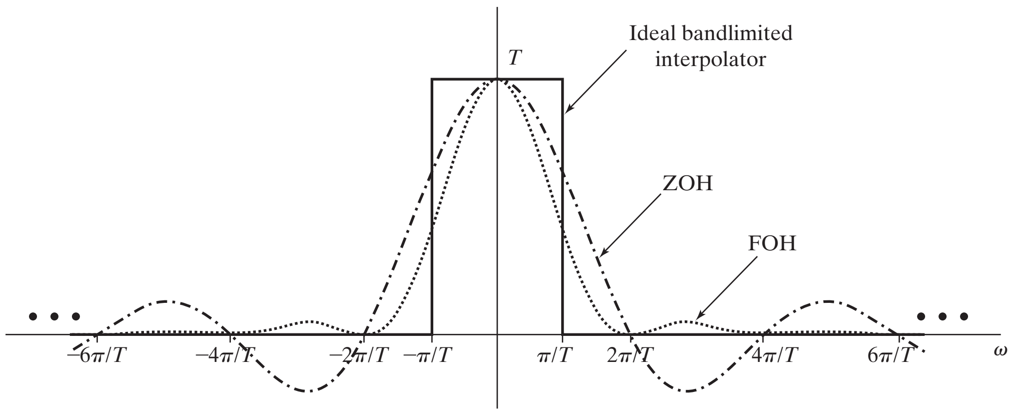


Figure 1.9 The Fourier transform amplitudes of the ideal bandlimited interpolator, the ZOH, and the FOH.

and that of the centered FOH pulse is

$$P_f(j\omega) = T\left(\frac{\sin(\omega T/2)}{\omega T/2}\right)^2 . \tag{1.100}$$

The Fourier transform amplitudes of the ideal bandlimited interpolator, the ZOH, and the FOH are shown in Figure 1.9.

1.6 FURTHER READING

As noted in the Preface, we assume a background in the foundations of signals and systems analysis in both continuous and discrete time. Chapters 1 and 2, which follow the development in [Op1] quite closely, are primarily intended to review and summarize basic concepts and establish notation. Computational explorations of this material are found in [Buc] and [McC]. Other texts on the basics of signals and systems include [Ch1], [Ha1], [Kwa], [La1], [Phi], and [Rob]. A rich set of perspectives is found in [Sie], which emphasizes continuous-time signals and systems. A somewhat more advanced development for discrete-time signals and systems is in [Op2], see also [Mit], [Ma1], [Pra] and [Pr1]. The geometric treatment in [Vet] exploits the view of signals as Hilbert-space vectors. Classic and fairly advanced books on signal analysis and Fourier transforms are [Bra], [Gui], [Pa1], [Pa2], [Pa3], all of which offer useful viewpoints. The treatment of Fourier theory in [Cha] is concise and illuminating.

Problems

Basic Problems

1.1. A simple physical model for the motion of a certain electric vehicle along a track is given by the following differential equation, with the position of the vehicle denoted by $y(t)$:

$$\frac{d^2y(t)}{dt^2} = -\left(\frac{c_f}{m}\right)\frac{dy(t)}{dt} - \left(\frac{c_b}{m}\right)\frac{dy(t)}{dt}x_b(t) + x_a(t)\,,$$

where $x_b(t)$ is the braking force applied to the wheels; $x_a(t)$ is the acceleration provided by the electric motor; m is the mass of the car; and c_f and c_b are frictional constants for the vehicle and brakes, respectively. Assume that we have the constraint $x_b \geq 0$, but that x_a can be positive or negative.

(a) Is the model linear? That is, do its nonzero solutions obey the superposition principle? Is the model time-invariant?

(b) How do your answers change if the braking force $x_b(t)$ is identically zero?

1.2. **(a)** Suppose the input signal to a stable LTI system with system function $H(s)$ is constant at some value α for all time t. What is the corresponding output at each t?

(b) Denote by $y(t)$ the output signal obtained from the system in (a) when the input to it is the signal $x(t) = t$ for all time. Now obtain two distinct expressions for the output corresponding to the input $t - \alpha$, where α is an arbitrary constant. *Hint*: Invoke the linearity and time invariance of the system, and use your result from (a). By choosing α appropriately, deduce that $y(t) = bt + y(0)$ for some constant b. Express b in terms of $H(s)$.

1.3. Indicate whether the systems below satisfy the following system properties: linearity, time invariance, causality, and BIBO stability.

(a) A system with input $x(t)$ and output $y(t)$, with input-output relation

$$y(t) = x^4(t), \quad -\infty < t < \infty\,.$$

(b) A system with input $x[n]$ and output $y[n]$, and input-output relation

$$y[n] = \begin{cases} 0 & n \leq 0 \\ y[n-1] + x[n] & n > 0\,. \end{cases}$$

(c) A system with input $x(t)$ and output $y(t)$, with input-output relation

$$y(t) = x(4t+3) \quad -\infty < t < \infty\,.$$

(d) A system with input $x(t)$ and output $y(t)$, with input-output relation

$$y(t) = \int_{-\infty}^{\infty} x(\tau)\,d\tau \quad -\infty < t < +\infty\,.$$

1.4. We are given a certain LTI system with impulse response $h_0(t)$, and are told that when the input is $x_0(t)$, the output $y_0(t)$ is the waveform shown in Figure P1.4.

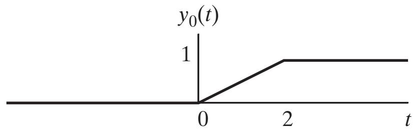


Figure P1.4

We are then given the following set of inputs $x(t)$ to LTI systems with the indicated impulse responses $h(t)$:

	Input $x(t)$	Impulse response $h(t)$
(a)	$x(t) = 2x_0(t)$	$h(t) = h_0(t)$
(b)	$x(t) = x_0(t) - x_0(t-2)$	$h(t) = h_0(t)$
(c)	$x(t) = x_0(t-2)$	$h(t) = h_0(t+1)$
(d)	$x(t) = x_0(-t)$	$h(t) = h_0(t)$
(e)	$x(t) = x_0(-t)$	$h(t) = h_0(-t)$
(f)	$x(t) = \frac{dx_0(t)}{dt}$	$h(t) = \frac{dh_0(t)}{dt}$

In each of these cases, determine whether or not we have enough information available to determine the output $y(t)$ uniquely. If it is possible to determine $y(t)$ uniquely, provide an analytical expression for it and a sketch of it. In those cases where you believe it is not possible to find $y(t)$ uniquely, see if you can prove that this is not possible.

1.5. **(a)** Consider an LTI system with input $x(t)$ and output $y(t)$ related through the equation

$$y(t) = \int_{-\infty}^{t} e^{-(t-\tau)} x(\tau - 2)\, d\tau \ .$$

What is the impulse response $h(t)$ for this system?

(b) Determine the response of this system when the input $x(t)$ is as shown in Figure P1.5-1.

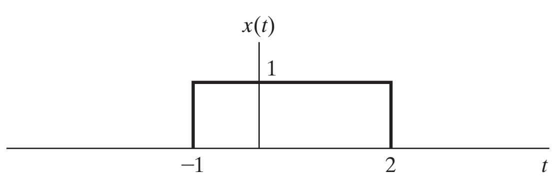


Figure P1.5-1

(c) Consider the interconnection of LTI systems shown in Figure P1.5-2. Here $h(t)$ is as in part (a). Determine the output $w(t)$ when the input $x(t)$ is the