DIGITAL FILTERS
Third Edition

R. W. HAMMING

DIGITAL FILTERS
Third Edition

R. W. HAMMING

Naval Postgraduate School
Monterey, California

PRENTICE HALL, ENGLEWOOD CLIFFS, NEW JERSEY 07632

LIBRARY OF CONGRESS
Library of Congress Cataloging-in-Publication Data

Hamming, R. W. (Richard Wesley)
Digital filters / R.W. Hamming. -- 3rd ed.
p. cm. -- (Prentice Hall signal processing series)
Bibliography: p.
Includes index.
ISBN 0-13-212812-8
1. Digital filters (Mathematics) I. Title. II. Series.
QA297.H26 1983
621.38'043--dc19 88-19598
CIP

Editorial/production supervision: Merrill Peterson
Cover design: Karen Stephens
Manufacturing buyer: Mary Noonan

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-212812-8

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon and Schuster Asia Pte Ltd , *Singapore*
Editora Prentice-Hall do Brasil, Ltda , *Rio de Janeiro*

Preface to the Third Edition

This edition retains the same basic approach of the earlier editions of stressing fundamentals; however, some changes have been made to reflect the fact that increasingly often a digital filter course is the first course in electrical engineering and the field of signal processing. To meet these needs two main changes have been made: (1) the inclusion of more material on the z-transform, which is often used in later courses (though the constant use of the formalism tends to obscure the ideas behind the manipulations), and (2) the inclusion of more examples and exercises. There are, of course, many minor changes to clarify and adapt the material to current uses.

In the years since I wrote the first edition I have become increasingly convinced of the need for a very elementary treatment of the subject of digital filters. The need for an elementary introduction comes from the fact that many of the people who most need the knowledge are not mathematically sophisticated and do not have an elaborate electrical engineering background. Thus this book assumes *only* a knowledge of the calculus and a smattering of statistics (which is reviewed in the text). It does not assume any electrical engineering background knowledge. Actually, experience seems to show that a prior knowledge of the corresponding theory of analog filters often causes more harm than good! Digital filtering is not simply converting from analog to digital filters; it is a fundamentally different way of thinking about the topic of signal processing, and many of the ideas and limitations of the analog method have no counterpart in the digital form.

The subject of digital filters is the natural introduction to the broad, fundamental field of signal processing. The power and basic simplicity of digital signal processing over the older analog is so great that whenever possible we are converting present analog systems to an equivalent digital form. But much more important, digital signaling allows fundamentally new things to be done easily. The availability of modern integrated-circuit chips, as well as micro- and minicomputers, has greatly expanded the application of digital filters.

Digital signals occur in many places. The telephone company is rapidly converting to the use of digital signals to represent the human voice. Even radio, television, and hi-fi sound systems are moving toward the all digital methods since they provide such superior fidelity and freedom from noise, as well as much more flexible signal processing. The space shots use digital signaling to transmit the information from the planets back to Earth, including the extremely detailed pictures (which were often processed digitally here on Earth to extract further information and to form alternate views of what was originally captured by the cameras in space). Most records of laboratory experiments are now recorded in digital form, from isolated measurements using a digital voltmeter to the automatic recording of entire sets of functions via a digital computer. Thus these signals are immediately ready for digital signal processing to extract the message that the experiment was designed to reveal. Economic data, from stock market prices and averages to the Cost of Living Index of the Bureau of Labor Statistics, occur only in digital form.

Digital filtering includes the processes of smoothing, predicting, differentiating, integrating, separation of signals, and removal of noise from a signal. Thus many people who do such things are actually using digital filters without realizing that they are; being unacquainted with the theory, they neither understand what they have done nor the possibilities of what they might have done. Computer people very often find themselves involved in filtering signals when they have had no appropriate training at all. Their needs are especially catered to in this book.

Because the same ideas arise in many fields there are many cross connections between the fields that can be exploited. Unfortunately each field seems to go its own way (while reinventing the wheel) and to develop its own jargon for exactly the same ideas that are used elsewhere. One goal of this revision is to expose and reduce this elaborate jargon equivalence from the various fields of application and to provide a unified approach to the whole field. We will adopt the simplest, most easily understood words to describe what is going on and exhibit lists of the equivalent words from related fields. We will also use only the simplest, most direct mathematical tools and shun fancy mathematics whenever possible.

This book concentrates on linear signal processing; the main exceptions are the examination of roundoff effects and a brief mention of Kalman filters, which adapt themselves to the signal they are receiving.

The fundamental tool of digital filtering is the frequency approach, which is based on the use of sines and cosines rather than on the use of polynomials (as is conventional in many fields such as numerical analysis and much of statistics). The frequency approach, which leads to the spectrum, has been the principal method of opening the black boxes of nature. Examples run from the early study of the structure of the atom (using spectral lines as the observations) through quantum mechanics (which arose from the study of the spectrum of black-body radiation) to the modern methods of studying a system (for purposes of modeling and control) via the spectrum of the output as it is related to the input.

There appears to be a deep emotional resistance to the frequency approach. And even electrical engineers who use it daily often have only a slight understanding of *why* they are using the eigenfunction approach and the role of the eigenvalues. In numerical analysis there is almost complete antipathy to the frequency approach, while in statistics there is a great fondness for polynomials (without ever examining the question of which set of functions is appropriate). This book shows clearly why the sines and cosines are the natural, the proper, the characteristic functions to use in many situations. It also approaches cautiously the usual traumatic experience (for most people) of going from the real sines and cosines to the complex exponentials with the mysterious $\sqrt{-1}$; their greater convenience in use eventually compensates for the initial troubles and provides more insight.

The text includes an accurate (but not excessively rigorous) introduction to the necessary mathematics. In each case the formal mathematics is postponed until the need for it is clearly seen. We are interested in presenting the *ideas* of the field and will generally not give the "best" methods for designing very complex filters; in an elementary course it is proper to give elementary, broadly applicable design methods, and then show how these can be refined to meet a very wide range of design criteria. Because it is an elementary text, references to advanced papers and books are of little use to the reader. Instead we refer to a few standard texts where more advanced material and references can be found. The references to these books are indicated in the text by [L,p], where L is the book label given at the end of this book, and p is the page(s) where it can be found. References [IEEE-1 and 2] give a complete bibliography for most topics that arise.

There is a deliberate repetition in the presentation of the material. Experience shows that the learner often becomes so involved in the immediate details of designing a filter that where and how the topic fits into the whole plan is lost. Furthermore, confusion often arises when the same ideas and mathematical tools are used in seemingly very different situations. It is also true that filters are designed to process data, but experience shows that the display of large sets of data that have been processed communicates very little to the beginner. Thus such plots are seldom given, even though the learner needs to be reminded that the ultimate test of a filter is how well it processes a signal, not how elegant the derivation is.

As always an author is deeply indebted to others, in this case to his many colleagues at Bell Laboratories. Special mention should go to Pro-

fessor J. W. Tukey (of Princeton University) and to J. F. Kaiser, who first taught him most of what is presented here. Thanks are also due to Roger Pinkham and the many students of the short courses who used the first two editions; their questions and reactions have been important in many places of this revision. They have also strengthened the author's belief in the basic rightness of giving as simple an approach as possible and of keeping rigorous mathematics in its proper place. Finally, thanks are due to the Naval Postgraduate School for providing an atmosphere suitable for thinking deeply about the problems of teaching.

R. W. Hamming

Contents

9 KAISER WINDOWS AND OPTIMIZATION 185

10 THE FINITE FOURIER SERIES 208

11 THE SPECTRUM 219

12 RECURSIVE FILTERS 230

1

Introduction

1.1 WHAT IS A DIGITAL FILTER?

In our current technical society we often measure a continuously varying quantity. Some examples include blood pressure, earthquake displacements, voltage from a voice signal in a telephone conversation, brightness of a star, population of a city, waves falling on a beach, and the probability of death. All these measurements vary with time; we regard them as functions of time: $u(t)$ in mathematical notation. And we may be concerned with blood pressure measurements from moment to moment or from year to year. Furthermore, we may be concerned with functions whose independent variable is not time, for example the number of particles that decay in a physics experiment as a function of the energy of the emitted particle. Usually these variables can be regarded as varying continuously (analog signals) even if, as with the population of a city, a bacterial colony, or the number of particles in the physics experiment, the number being measured must change by unit amounts.

For technical reasons, instead of the signal $u(t)$, we usually record *equally spaced samples* u_n of the function $u(t)$. The famous *sampling theorem*, which will be discussed in Chapter 8, gives the conditions on the signal that justify this sampling process. Moreover, when the samples are taken they are not recorded with infinite precision but are rounded off (sometimes chopped off) to comparatively few digits (see Figure 1.1-1). This procedure is often called *quantizing* the samples. It is these quantized samples that are available for the processing that we do. We do the processing in order to understand what the function samples u_n reveal about the underlying phenomena that gave rise to the observations, and digital filters are the main processing tool.

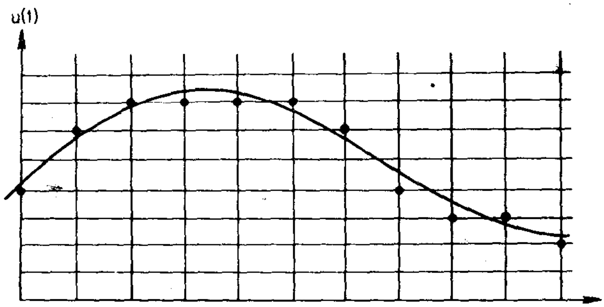

FIGURE 1.1-1 SAMPLING AND QUANTIZATION OF A SIGNAL

It is necessary to emphasize that the samples are *assumed* to be equally spaced; any error or *noise* is in the measurements u_n. Fortunately, this assumption is approximately true in most applications.

Suppose that the sequence of numbers $\{u_n\}$ is such a set of equally spaced measurements of some quantity $u(t)$, where n is an integer and t is a continuous variable. Typically, t represents time, but not necessarily so. We are using the notation $u_n = u(n)$. The simplest kinds of filters are the *nonrecursive filters*; they are defined by the linear formula

$$y_n = \sum_{k=-\infty}^{\infty} c_k u_{n-k} \tag{1.1-1}$$

The coefficients c_k are the constants of the filter, the u_{n-k} are the input data, and the y_n are the outputs. Figure 1.1-2 shows how this formula is computed. Imagine two strips of paper. On the first strip, written one below the other, are the data values u_{n-k}. On the second strip, with the values written in the *reverse direction* (from bottom to top), are the filter coefficients c_k. The zero subscript of one is opposite the n subscript value of the other (either way). The output y_n is the sum of all the products $c_k u_{n-k}$. Having computed one value, one strip, say the coefficient strip, is moved one space down, and the new set of products is computed to give the new output y_{n+1}. Each output is the result of adding all the products formed from the proper displacement between the two zero-subscripted terms. In the computer, of course, it is the data that is "run past" the coefficient array $\{c_k\}$.

This process is basic and is called a *convolution* of the data with the coefficients. It does not matter which strip is written in the reverse order; the result is the same. So the convolution of u_n with the coefficients c_k is the same as the convolution of the coefficients c_k with the data u_n.

In practice, the number of products we can handle must be finite. It is usual to assume that the length of the run of nonzero coefficients c_k is much shorter than is the run of data y_n. Once in a while it is useful to regard the

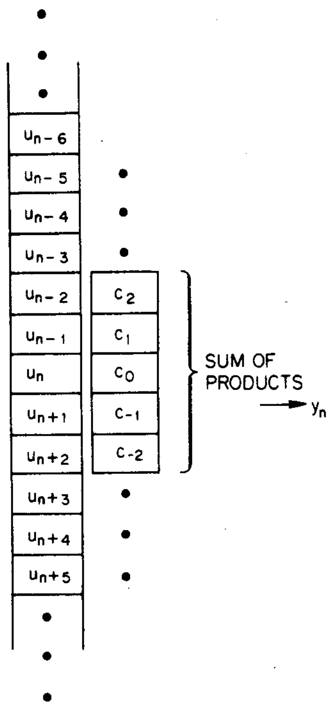

FIGURE 1.1-2 A NONRECURSIVE DIGITAL FILTER

c_k coefficients as part of an infinite array with many zero coefficients, but it is usually preferable to think of the array $\{c_k\}$ as being finite and to ignore the zero terms beyond the end of the array. Equation (1.1-1) becomes, therefore,

$$y_n = \sum_{k=-N}^{N} c_k u_{n-k} \tag{1.1-2}$$

Thus the second strip (of coefficients c_k) in Figure 1.1-2 is comparatively shorter than is the first strip (of data u_n).

Various special cases of this formula occur frequently and should be familiar to most readers. Indeed, such formulas are so commonplace that a book could be devoted to their listing. In the case of five nonzero coefficients c_k, where all the coefficients that are not zero have the same value, we have the familiar smoothing by 5s formula (derived in Section 3.2)

$$y_n = \tfrac{1}{5}(u_{n-2} + u_{n-1} + u_n + u_{n+1} + u_{n+2}) \tag{1.1-3}$$

Another example is the least-squares smoothing formula derived by passing a least-squares cubic through five equally spaced values u_n and using the value of the cubic at the midpoint as the smoothed value. The formula for this smoothed value (which will be derived in Section 3.3) is

$$y_n = \frac{1}{35}(-3u_{n-2} + 12u_{n-1} + 17u_n + 12u_{n+1} - 3u_{n+2}) \qquad (1.1\text{-}4)$$

Many other formulas, such as those for predicting stock market prices, as well as other time series, also are nonrecursive filters.

Nonrecursive filters occur in many different fields and, as a result, have acquired many different names. Among the disguises are the following:

Finite impulse response filter

FIR filter

Transversal filter

Tapped delay line filter

Moving average filter

We shall use the name *nonrecursive* as it is the simplest to understand from its name, and it contrasts with the name *recursive filter*, which we will soon introduce.

The concept of a *window* is perhaps the most confusing concept in the whole subject, so we now introduce it in these simple cases. We can think of the preceding formulas as if we were looking at the data u_{n-k} through a *window of coefficients* c_k (see Figure 1.1-3). As we slide the strip of coefficients along the data, we see the data in the form of the output y_n, which

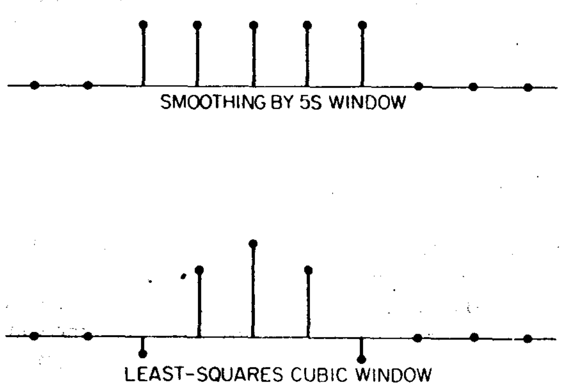

FIGURE 1.1-3 WINDOWS

is the running weighted average of the original data u_n. It is as if we saw the data through a translucent (not transparent) window where the window was tinted according to the coefficients c_k. In the smoothing by 5s, all data values get through the translucent window with the same amount, $\frac{1}{5}$; in the second example they come through the window with varying weights. (Don't let any negative weights bother you, since we are merely using a manner of speaking when we use the words "translucent window.")

When we use not only data values to compute the output values y_n but also use other values of the output, we have a formula of the form

$$y_n = \sum_{-\infty}^{\infty} c_k u_{n-k} + \sum_{-\infty}^{\infty} d_k y_{n-k}$$

where both the c_k and the d_k are constants. In this case it is usual to limit the range of nonzero coefficients to current and past values of the data u_n and to only past values of the output y_n. Furthermore, again the number of products that can be computed in practice must be finite. Thus the formula is usually written in the form

$$y_n = \sum_{0}^{N} c_k u_{n-k} + \sum_{1}^{M} d_k y_{n-k} \qquad (1.1\text{-}5)$$

where there may be some zero coefficients. These are called *recursive* filters (see Figure 1.1-4). Some equivalent names follow:

Infinite impulse response filter

IIR filter

Ladder filter

Lattice filter

Wave digital filter

Autoregressive moving average filter

ARMA filter

Autoregressive integrated moving average filter

ARIMA filter

We shall use the name *recursive* filter. A recursive digital filter is simply a linear difference equation with constant coefficients and nothing more; in practice it may be realized by a short program on a general purpose digital computer or by a special purpose integrated circuit chip.