Digital Signal Processing and Digital Filter Design (Draft)

By:

C. Sidney Burrus

Digital Signal Processing and Digital Filter Design (Draft)

By:
C. Sidney Burrus

Online:
< http://cnx.org/content/col10598/1.6/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: November 17, 2012
PDF generated: November 17, 2012
For copyright and attribution information for the modules contained in this collection, see p. 310.

Table of Contents

Preface: Digital Signal Processing and Digital Filter Design[1]

Digital signal processing (DSP) has existed as long as quantitative calculations have been systematically applied to data in Science, Social Science, and Technology. The set of activities started out as a collection of ideas and techniques in very different applications. Around 1965, when the fast Fourier transform (FFT) was rediscovered, DSP was extracted from its applications and became a single academic and professional discipline to be developed as far as possible.

One of the earliest books on DSP was by Gold and Rader [125], written in 1968, although there had been earlier books on sampled data control and time series analysis, and chapters in books on computer applications. In the late 60's and early 70's there was an explosion of activity in both the theory and application of DSP. As the area was beginning to mature, two very important books on DSP were published in 1975, one by Oppenheim and Schafer [225] and the other by Rabiner and Gold [284]. These three books dominated the early courses in universities and self study in industry.

The early applications of DSP were in the defense, oil, and medical industries. They were the ones who needed and could afford the expensive but higher quality processing that digital techniques offered over analog signal processing. However, as the theory developed more efficient algorithms, as computers became more powerful and cheaper, and finally, as DSP chips became commodity items (e.g. the Texas Instruments TMS-320 series) DSP moved into a variety of commercial applications and the current digitization of communications began. The applications are now

[1] This content is available online at <http://cnx.org/content/m16880/1.2/>.

everywhere. They are tele-communications, seismic signal processing, radar and sonar signal processing, speech and music signal processing, image and picture processing, entertainment signal processing, financial data signal processing, medical signal processing, nondestructive testing, factory floor monitoring, simulation, visualization, virtual reality, robotics, and control. DSP chips are found in virtually all cell phones, digital cameras, high-end stereo systems, MP3 players, DVD players, cars, toys, the "Segway", and many other digital systems.

In a modern curriculum, DSP has moved from a specialized graduate course down to a general undergraduate course, and, in some cases, to the introductory freshman or sophomore EE course [198]. An exciting project is experimenting with teaching DSP in high schools and in colleges to non-technical majors [237].

Our reason for writing this book and adding to the already long list of DSP books is to cover the new results in digital filter design that have become available in the last 10 to 20 years and to make these results available on line in Connexions as well as print. Digital filters are important parts of a large number of systems and processes. In many cases, the use of modern optimal design methods allows the use of a less expensive DSP chip for a particular application or obtaining higher performance with existing hardware. The book should be useful in an introductory course if the students have had a course on discrete-time systems. It can be used in a second DSP course on filter design or used for self-study or reference in industry.

We first cover the optimal design of Finite Impulse Response (FIR) filters using a least squared error, a maximally flat, and a Chebyshev criterion. A feature of the book is covering finite impulse response (FIR) filter design before infinite impulse response (IIR) filter design. This reflects modern practice and new filter design algorithms. The FIR filter design chapter contains new methods on constrained optimization, mixed optimization criteria, and modifications to the basic Parks-McClellan algorithm that are very useful. Design programs are given in MatLab and FORTRAN.

A brief chapter on structures and implementation presents block processing for both FIR and IIR filters, distributed arithmetic structures for multiplierless implementation, and multirate systems for filter banks and wavelets. This is presented as a generalization to sampling and to periodically time-varying systems. The bifrequency map gives a clearer explanation of aliasing and how to control it.

The basic notes that were developed into this book have evolved over 35 years of teaching and conducting research in DSP at Rice, Erlangen, and MIT. They contain the results of research on filters and algorithms done at those universities and other universities and industries around the world. The book tries to give not only the different methods and approaches, but also reasons and intuition for choosing one method over another. It should be interesting to both the university student and the industrial practitioner.

We want to acknowledge with gratitude the long time support of Texas Instruments, Inc., the National Science Foundation, National Instruments, Inc. and the MathWorks, Inc. as well as the support of the Maxfield and Oshman families. We also want to thank our long-time colleagues Tom Parks, Hans Schuessler, Jim McClellan, Al Oppenheim, Sanjit Mitra, Ivan Selesnick, Doug Jones, Don Johnson, Leland Jackson, Rich Baraniuk, and our graduate students over 30 years from whom we have learned much and with whom we have argued often, particularly, Selesnick, Gopinath, Soewito, and Vargas. We also owe much to the IEEE Signal Processing Society and to Rice University for environments to learn, teach, create, and collaborate. Much of the results in DSP was supported directly or indirectly by the NSF, most recently NSF grant EEC-0538934 in the Partnerships for Innovation program working with National Instruments, Inc.

We particularly thank Texas Instruments and Prentice Hall for returning the copyrights to me so that part of the material in **DFT/FFT and Convolution Algorithms**[58], **Design of Digital Filters**[245], and "Efficient Fourier Transform and Convolution Algorithms" in **Advanced Topics in Signal Processing**[44] could be included here under the Creative Commons Attribution copyright. I also appreciate IEEE policy that allows parts of my papers to be included here.

A rather long list of references is included to point to more background, to more advanced theory, and to applications. A book of Matlab DSP exercises that could be used with this book has been published through Prentice Hall [56], [199]. Some Matlab programs are included to aid in understanding the design algorithms and to actually design filters. LabView from National Instruments is a very useful tool to both learn with and use in application. All of the material in these notes is being put into "Connexions" [22] which is a modern web-based open-content information system www.cnx.org. Further information is available on our web site at www.dsp.rice.edu with links to other related work. We thank Richard Baraniuk, Don Johnson, Ray Wagner, Daniel Williamson, and Marcia Horton for their help.

This version of the book is a draft and will continue to evolve under Connexions. A companion FFT book is being written and is also available in Connexions and print form. All of these two books are in the repository of Connexions and, therefore, available to anyone free to use, reuse, modify, etc. as long as attribution is given.

C. Sidney Burrus

Houston, Texas

June 2008

Chapter 1

Signals and Signal Processing Systems

1.1 Continuous-Time Signals[1]

Signals occur in a wide range of physical phenomenon. They might be human speech, blood pressure variations with time, seismic waves, radar and sonar signals, pictures or images, stress and strain signals in a building structure, stock market prices, a city's population, or temperature across a plate. These signals are often modeled or represented by a real or complex valued mathematical function of one or more variables. For example, speech is modeled by a function representing air pressure varying with time. The function is acting as a mathematical analogy to the speech signal and, therefore, is called an **analog** signal. For these signals, the independent variable is time and it changes continuously so that the term **continuous-time** signal is also used. In our discussion, we talk of the mathematical function as the signal even though it is really a model or representation of the physical signal.

The description of signals in terms of their sinusoidal frequency content has proven to be one of the most powerful tools of continuous and discrete-time signal description, analysis, and processing. For that reason, we will start the discussion of signals with a development of Fourier transform methods. We will first review the continuous-time methods of the Fourier series (FS), the Fourier transform or integral (FT), and the Laplace transform (LT). Next the discrete-time methods will be developed in more detail with the discrete Fourier transform (DFT) applied to finite

[1]This content is available online at <http://cnx.org/content/m16920/1.2/>.

length signals followed by the discrete-time Fourier transform (DTFT) for infinitely long signals and ending with the Z-transform which allows the powerful tools of complex variable theory to be applied.

More recently, a new tool has been developed for the analysis of signals. Wavelets and wavelet transforms [150], [63], [92], [380], [347] are another more flexible expansion system that also can describe continuous and discrete-time, finite or infinite duration signals. We will very briefly introduce the ideas behind wavelet-based signal analysis.

1.1.1 The Fourier Series

The problem of expanding a finite length signal in a trigonometric series was posed and studied in the late 1700's by renowned mathematicians such as Bernoulli, d'Alembert, Euler, Lagrange, and Gauss. Indeed, what we now call the Fourier series and the formulas for the coefficients were used by Euler in 1780. However, it was the presentation in 1807 and the paper in 1822 by Fourier stating that an arbitrary function could be represented by a series of sines and cosines that brought the problem to everyone's attention and started serious theoretical investigations and practical applications that continue to this day [147], [69], [165], [164], [116], [223]. The theoretical work has been at the center of analysis and the practical applications have been of major significance in virtually every field of quantitative science and technology. For these reasons and others, the Fourier series is worth our serious attention in a study of signal processing.

1.1.1.1 Definition of the Fourier Series

We assume that the signal $x(t)$ to be analyzed is well described by a real or complex valued function of a real variable t defined over a finite interval $\{0 \le t \le T\}$. The trigonometric series expansion of $x(t)$ is given by

$$x(t) = \frac{a(0)}{2} + \sum_{k=1}^{\infty} a(k)\, cos\left(\frac{2\pi}{T}kt\right) + b(k)\, sin\left(\frac{2\pi}{T}kt\right). \qquad (1.1)$$

where $x_k(t) = cos(2\pi kt/T)$ and $y_k(t) = sin(2\pi kt/T)$ are the basis functions for the expansion. The energy or power in an electrical, mechanical, etc. system is a function of the square of voltage, current, velocity, pressure, etc. For this reason, the natural setting for a representation of signals is the Hilbert space of $L^2[0,T]$. This modern formulation of the problem is developed in [104], [165]. The sinusoidal basis functions in the trigonometric expansion form a complete orthogonal set in $L^2[0,T]$. The orthogonality is easily seen from inner products

$$\left(cos\left(\tfrac{2\pi}{T}kt\right)\ ,\ cos\left(\tfrac{2\pi}{T}\ell t\right)\right) = \int_0^T \left(cos\left(\tfrac{2\pi}{T}kt\right)\ cos\left(\tfrac{2\pi}{T}\ell t\right)\right)\ dt = \delta(k-\ell) \tag{1.2}$$

and

$$\left(cos\left(\frac{2\pi}{T}kt\right)\ ,\ sin\left(\frac{2\pi}{T}\ell t\right)\right) = \int_0^T \left(cos\left(\frac{2\pi}{T}kt\right)\ sin\left(\frac{2\pi}{T}\ell t\right)\right)\ dt = 0 \tag{1.3}$$

where $\delta(t)$ is the Kronecker delta function with $\delta(0)=1$ and $\delta(k\neq 0)=0$. Because of this, the kth coefficients in the series can be found by taking the inner product of $x(t)$ with the kth basis functions. This gives for the coefficients

$$a(k) = \frac{2}{T}\int_0^T x(t)\,cos\left(\frac{2\pi}{T}kt\right)dt \tag{1.4}$$

and

$$b(k) = \frac{2}{T}\int_0^T x(t)\,sin\left(\frac{2\pi}{T}kt\right)dt \tag{1.5}$$

where T is the time interval of interest or the period of a periodic signal. Because of the orthogonality of the basis functions, a finite Fourier series formed by truncating the infinite series is an optimal least squared error approximation to $x(t)$. If the finite series is defined by

$$\hat{x}(t) = \frac{a(0)}{2} + \sum_{k=1}^{N} a(k)\,cos\left(\frac{2\pi}{T}kt\right) + b(k)\,sin\left(\frac{2\pi}{T}kt\right), \tag{1.6}$$

the squared error is

$$\epsilon = \frac{1}{T}\int_0^T |x(t) - \hat{x}(t)|^2\,dt \tag{1.7}$$

which is minimized over all $a(k)$ and $b(k)$ by (1.4) and (1.5). This is an extraordinarily important property.

It follows that if $x(t)\in L^2[0,T]$, then the series converges to $x(t)$ in the sense that $\epsilon\to 0$ as $N\to\infty$[104], [165]. The question of point-wise convergence is more difficult. A sufficient condition that is adequate for most application states: If $f(x)$ is bounded, is piece-wise continuous, and has no more than a finite number of maxima over an interval, the Fourier series converges point-wise to $f(x)$ at all points of continuity and to the arithmetic mean at points of discontinuities. If $f(x)$ is continuous, the series converges uniformly at all points [165], [147], [69].

A useful condition [104], [165] states that if $x(t)$ and its derivatives through the qth derivative are defined and have bounded variation, the Fourier coefficients $a(k)$ and $b(k)$ asymptotically drop off at least as fast as $\frac{1}{k^{q+1}}$ as $k \to \infty$. This ties global rates of convergence of the coefficients to local smoothness conditions of the function.

The form of the Fourier series using both sines and cosines makes determination of the peak value or of the location of a particular frequency term difficult. A different form that explicitly gives the peak value of the sinusoid of that frequency and the location or phase shift of that sinusoid is given by

$$x(t) = \frac{d(0)}{2} + \sum_{k=1}^{\infty} d(k)\, cos\left(\frac{2\pi}{T}kt + \theta(k)\right) \tag{1.8}$$

and, using Euler's relation and the usual electrical engineering notation of $j = \sqrt{-1}$,

$$e^{jx} = cos(x) + j sin(x), \tag{1.9}$$

the complex exponential form is obtained as

$$x(t) = \sum_{k=-\infty}^{\infty} c(k)\; e^{j\frac{2\pi}{T}kt} \tag{1.10}$$

where

$$c(k) = a(k) + j\, b(k). \tag{1.11}$$

The coefficient equation is

$$c(k) = \frac{1}{T}\int_0^T x(t)\; e^{-j\frac{2\pi}{T}kt} dt \tag{1.12}$$

The coefficients in these three forms are related by

$$|d|^2 = |c|^2 = a^2 + b^2 \tag{1.13}$$

and

$$\theta = arg\{c\} = tan^{-1}\left(\frac{b}{a}\right) \tag{1.14}$$

It is easier to evaluate a signal in terms of $c(k)$ or $d(k)$ and $\theta(k)$ than in terms of $a(k)$ and $b(k)$. The first two are polar representation of a complex value and the last is rectangular. The exponential form is easier to work with mathematically.

Although the function to be expanded is defined only over a specific finite region, the series converges to a function that is defined over the real line and is periodic. It is equal to the original function over the region of definition and is a periodic extension outside of the region. Indeed, one could artificially extend the given function at the outset and then the expansion would converge everywhere.

1.1.1.2 A Geometric View

It can be very helpful to develop a geometric view of the Fourier series where $x(t)$ is considered to be a vector and the basis functions are the coordinate or basis vectors. The coefficients become the projections of $x(t)$ on the coordinates. The ideas of a measure of distance, size, and orthogonality are important and the definition of error is easy to picture. This is done in [104], [165], [390] using Hilbert space methods.

1.1.1.3 Properties of the Fourier Series

The properties of the Fourier series are important in applying it to signal analysis and to interpreting it. The main properties are given here using the notation that the Fourier series of a real valued function $x(t)$ over $\{0 \leq t \leq T\}$ is given by $\mathcal{F}\{x(t)\} = c(k)$ and $\tilde{x}(t)$ denotes the periodic extensions of $x(t)$.

1. Linear: $\mathcal{F}\{x+y\} = \mathcal{F}\{x\} + \mathcal{F}\{y\}$
   Idea of superposition. Also scalability: $\mathcal{F}\{ax\} = a\mathcal{F}\{x\}$
2. Extensions of $x(t)$: $\tilde{x}(t) = \tilde{x}(t+T)$
   $\tilde{x}(t)$ is periodic.
3. Even and Odd Parts: $x(t) = u(t) + jv(t)$ and $C(k) = A(k) + jB(k) = |C(k)|\, e^{j\theta(k)}$

u	v	A	B	$\|C\|$	θ
even	0	even	0	even	0
odd	0	0	odd	even	0
0	even	0	even	even	$\pi/2$
0	odd	odd	0	even	$\pi/2$

Table 1.1

4. Convolution: If continuous cyclic convolution is defined by

$$y(t) \; = \; h(t) \circ x(t) \; = \; \int_0^T \tilde{h}(t-\tau) \; \tilde{x}(\tau) \; d\tau \qquad (1.15)$$

then $\mathcal{F}\{h(t) \circ x(t)\} \; = \; \mathcal{F}\{h(t)\} \, \mathcal{F}\{x(t)\}$

5. Multiplication: If discrete convolution is defined by

$$e(n) \; = \; d(n) * c(n) \; = \; \sum_{m=-\infty}^{\infty} d(m) \; c(n-m) \qquad (1.16)$$

then $\mathcal{F}\{h(t) \; x(t)\} \; = \; \mathcal{F}\{h(t)\} * \mathcal{F}\{x(t)\}$
This property is the inverse of property 4 (list, p. 9) and vice versa.

6. Parseval: $\frac{1}{T} \int_0^T |x(t)|^2 dt \; = \; \sum_{k=-\infty}^{\infty} |C(k)|^2$
This property says the energy calculated in the time domain is the same as that calculated in the frequency (or Fourier) domain.
7. Shift: $\mathcal{F}\{\tilde{x}(t-t_0)\} \; = \; C(k) \; e^{-j2\pi t_0 k/T}$
A shift in the time domain results in a linear phase shift in the frequency domain.
8. Modulate: $\mathcal{F}\{x(t) \; e^{j2\pi Kt/T}\} \; = \; C(k-K)$
Modulation in the time domain results in a shift in the frequency domain. This property is the inverse of property 7.
9. Orthogonality of basis functions:

$$\int_0^T e^{-j2\pi mt/T} \, e^{j2\pi nt/T} \, dt \; = \; T \, \delta(n-m) \; = \; \left\{ \begin{array}{ll} T & \text{if } n = m \\ 0 & \text{if } n \neq m. \end{array} \right. \qquad (1.17)$$

Orthogonality allows the calculation of coefficients using inner products in (1.4) and (1.5). It also allows Parseval's Theorem in property 6 (list, p. 10). A relaxed version of orthogonality is called “tight frames" and is important in over-specified systems, especially in wavelets.

1.1.1.4 Examples

- An example of the Fourier series is the expansion of a square wave signal with period 2π. The expansion is

$$x(t) \; = \; \frac{4}{\pi} \left[sin(t) + \frac{1}{3} sin(3t) + \frac{1}{5} sin(5t) \cdots \right]. \qquad (1.18)$$

Because $x(t)$ is odd, there are no cosine terms (all $a(k) = 0$) and, because of its symmetries, there are no even harmonics (even k terms are zero). The function is well defined and bounded; its derivative is not, therefore, the coefficients drop off as $\frac{1}{k}$.

- A second example is a triangle wave of period 2π. This is a continuous function where the square wave was not. The expansion of the triangle wave is

$$x(t) = \frac{4}{\pi}\left[sin(t) - \frac{1}{3^2}sin(3t) + \frac{1}{5^2}sin(5t) + \cdots\right]. \tag{1.19}$$

Here the coefficients drop off as $\frac{1}{k^2}$ since the function and its first derivative exist and are bounded.

Note the derivative of a triangle wave is a square wave. Examine the series coefficients to see this. There are many books and web sites on the Fourier series that give insight through examples and demos.

1.1.1.5 Theorems on the Fourier Series

Four of the most important theorems in the theory of Fourier analysis are the inversion theorem, the convolution theorem, the differentiation theorem, and Parseval's theorem [71].

- The inversion theorem is the truth of the transform pair given in (1.1), (1.4), and (1.5)..
- The convolution theorem is property 4 (list, p. 9).
- The differentiation theorem says that the transform of the derivative of a function is $j\omega$ times the transform of the function.
- Parseval's theorem is given in property 6 (list, p. 10).

All of these are based on the orthogonality of the basis function of the Fourier series and integral and all require knowledge of the convergence of the sums and integrals. The practical and theoretical use of Fourier analysis is greatly expanded if use is made of distributions or generalized functions (e.g. Dirac delta functions, $\delta(t)$) [239], [32]. Because energy is an important measure of a function in signal processing applications, the Hilbert space of L^2 functions is a proper setting for the basic theory and a geometric view can be especially useful [104], [71].

The following theorems and results concern the existence and convergence of the Fourier series and the discrete-time Fourier transform [226]. Details, discussions and proofs can be found in the cited references.