FOURIER TRANSFORM

The importance of Fourier transforms or, more generally, harmonic analysis to electrical engineering and telecommunications can hardly be overstated. It is a prime example of how a mathematical theory changed our lives and made possible the ubiquitous communication we are now accustomed to. It has had a decisive influence on other areas as well, most notably physics and quantum mechanics. Interest in harmonic analysis has not abated. The field has been rich with an everchanging array of important new problems, and a recent addition is the "wavelet" theory. In view of the central role it plays for our understanding of nature and its technical applications, it is safe to state that harmonic analysis will remain a crucial player in the coming century.

Fourier transform theory treats the representation of a signal as the sum or integral of sine waves, called the harmonic content of the signal. Since many signal processing devices or media, including our own atmosphere, are linear, they will propagate sine waves without altering their frequency, as sine waves, and will transfer a weighted sum of them to another such weighted sum. Harmonic analysis allows for efficient computations on signals and systems, and even for efficient arithmetic via the FFT—fast Fourier transform—and its derivatives.

HISTORICAL NOTES

In their book (1), Dym and McKean place the first occurrence of modern harmonic analysis in the work of d'Alembert on oscillations of a violin string. Further contributions in the eighteenth century were made by Euler, Bernouilli, and Lagrange. Fourier's work dates back to the early nineteenth century in which he laid the groundwork for the proof that smooth functions could be expanded in a trigonometric sum (2). There were many further contributions in the nineteenth century, and an account is given in Refs. 3 and 4.

However, the big breakthroughs came in the twentieth century, in several directions. First there was the work of Stieltjes and Lebesgue on generalized integrals and the introduction of L_1 and L_2 spaces, which also gave precise meanings to Fourier transform integrals. From them the very attractive Fourier isomorphism theory could be deduced. Hardy space theory ensued, and it provided the proper setting to study linear time invariant and stochastic systems, especially from the point of view of energetic behavior. Important problems such as system approximation, inversion, selective filtering, and linear least-squares approximation of stochastic variables could be solved in this framework. The interplay between Fourier theory and complex function analysis led to fundamental contributions by Hardy, Littlewood, Wiener, and others.

As nice as it was, function theory, even in the Lebesgue sense, did not answer all the questions of mathematical and engineering interest; in particular, it was not able to handle "generalized functions" which do, indeed, have interesting physical counterparts and which allow for considerable engineering simplifications. This led to a number of more or less rigorous attempts at breaking out first of the L_2 stranglehold and next out of the function theory context altogether. The most successful was the Laplace transform—an elegant and reasonably painless extension of Fourier theory, but with great applications for the solution of partial differential equations. Generalized functions were not put on a firm theoretical basis until the introduction of "distributions" by Laurent Schwartz in the mid-1950s. Although the distribution theory does involve advanced mathematical concepts, its basic principles can easily be grasped and used by practitioners of signal and system engineering, because they do have some physical appeal and are reasonably simple to interpret in a concrete situation.

On another front, the arrival of electronic computing greatly enhanced the interest in the computation of discrete Fourier transforms, for various practical reasons such as signal identification and filtering. But here something fundamentally new happened; It was noticed, first by Good and Thomas in the early 1950s, and in 1965 by Cooley and Tuckey, that the Fourier transform could be calculated much more efficiently than its direct definition would suggest. This led to a first version of the FFT, many more would follow. Soon after the appearance of the Cooley–Tuckey article, Stockman showed that this would lead to an efficient convolution algorithm as well, and hence to efficient algorithms to solve an array of problems in which convolution plays a central role such as multiplication of numbers or solving special kinds of systems of structured matrix equations—for example, systems with Toeplitz and Hankel matrices. The fast Fourier transform and its many improvements have become a prime example of efficient algorithmic thinking for the computer age.

THE BASIC FRAMEWORK FOR HARMONIC ANALYSIS: HILBERT SPACES

Although we might retrace the steps the researchers of the nineteenth century (like Fourier) undertook in the discovery of what we now call Fourier theory, we will not do so, but put ourselves squarely in the context of the modern mathematical theory of Hilbert spaces of the ℓ_2 or L_2 types, one of the natural habitats for Fourier theory of use both in electrical engineering and fundamental physics. We shall explore another habitat, the theory of generalized functions, of importance to communication theory, in a later section. In this section we give a brief introduction to Hilbert space theory, just enough to provide the necessary background for the main definitions.

From an engineering point of view, Hilbert spaces are important as soon as energy considerations come into play. For example, suppose that $v(t)$ is a voltage observed across a constant resistance R over a time interval $t \in [0, T]$. The energy dissipated by the resistor is given by the expression $\mathscr{E}(t) = (1/R)\int_0^T |v(t)|^2\, dt$. The integral of a time-dependent quantity square over a time interval is characteristic for expressions involving energy. Hilbert space theory studies such functions when their "size" or norm is measured in terms of such an energy integral. Of interest is in particular how a function $v(t)$ can be approximated by another such function with some special characteristics; clever approximation theory with relevant approximating functions is one of the main goals of harmonic analysis. To be more specific, assume that $v(t)$ is continuous, and that it is zero at both end points of the interval: $v(0) = 0$ and $v(T) = 0$. Consider now some special functions on the interval $[0, T]$ which meet the boundary values. Let n be an integer, and define $s_n(t) = \sin[\pi n(t/T)]$. We may wonder whether we can approximate our original $v(t)$ with a sum $\sum_{n=0}^{N} a_n s_n(t)$ in which a_n are fixed coefficients and N is a small number. It turns out that, indeed, this is often the case. Especially when $v(t)$ does not vary too quickly, N can be chosen very small; and in effect, $v(t)$ is represented by a small collection of coefficients $\{a_n\}_0^N$, called its relevant Fourier coefficients, and the terms $a_n s_n(t)$ are its constituting "harmonics." The famous "direct cosine transform" (DCT) used in the MPEG-2 standard for coding of video images in multimedia technology is but an elaboration of this principle!

Let us focus for a moment on the square root of the integral defined earlier and assign to it a special symbol:

$$\|v\|_2 = \left\{ \int_0^T |v(t)|^2\, dt \right\}^{1/2} \tag{1}$$

We call it a *norm* for v. Such a norm has the following characteristic properties:

1. If a is any number, then $\|av\| = |a|\|v\|$.
2. If v_1 and v_2 are two time functions on the interval $[0, T]$, then

$$\|v_1 + v_2\| \le \|v_1\| + \|v_2\| \tag{2}$$

   with equality when, and only when, the v_1 and v_2 are scaled versions of the same function (*the triangle inequality*).
3. For most classes of "reasonable" functions (we must make exception for generalized functions defined in a further section), $v(t) = 0$ if and only if $\|v\| = 0$ [$v(t)$ may have unessential "point defects," and the correct expression is: $v(t)$ is zero *almost everywhere* (a.e.)].

The quadratic norm has another remarkable property, namely that it can be derived from (and also generates) an inner product. Let us consider two time functions $v_1(t)$ and $v_2(t)$; and for good generality, let us assume that they can take complex values (real-valued functions are then a special case). We take $j = \sqrt{-1}$ as the complex number with magnitude 1 and phase $\pi/2$, and we denote complex conjugation by a superbar: $\bar{a}$ is the complex conjugate of a. The inner product between two functions v and w is defined as

$$(v, w) = \int_0^T \overline{w(t)} v(t)\, dt \tag{3}$$

Hence we see that

$$\|v\|_2^2 = (v, v) \tag{4}$$

The inner product has a number of properties which are worth recording:

1. For any complex number a: $(av, w) = a(v, w)$.
2. $(v, w) = \overline{(w, v)}$.
3. $(v_1 + v_2, w) = (v_1, w) + (v_1, w)$.
4. $(v, v)^{1/2}$ is a norm.

We say that because of properties 2 and 3 the inner product is "sesquilinear;" it is linear in the first argument, but only conjugate linear in the second. Although we require that an inner product actually defines a norm, we can also recover it from a norm via the formula

$$(v, w) = \tfrac{1}{4}[\|v + w\|^2 - \|v - w\|^2 + i\|v + iw\|^2 - i\|v - iw\|^2] \tag{5}$$

One may wonder whether every norm gives rise to an inner product. The answer is an emphatic "no;" only those norms that satisfy the parallellogram rule (henceforth we drop the subscript 2 from the norm size if is clear from the context)

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2) \tag{6}$$

do. Typically, these norms will be quadratic.

The notion of a Hilbert space starts out from a set of elements such as $\{v(t)\}$ on which addition and multiplication with scalars is well-defined (the set is a *linear space*), takes the existence and properties of an inner product and corresponding norm for granted, and focuses on the possibilities of approximation. We say that the space forms a Hilbert space $\mathcal{H}$ if, in addition to the previous properties, a property called *completeness* (which we now introduce) holds. We say that a collection $\{v_n\}$ of members of $\mathcal{H}$ indexed by integers n has the *Cauchy property* if for any arbitrarily small positive number ϵ there exists an integer N_ϵ such that $\|v_n - v_m\| < \epsilon$ as soon as both n and m are greater than N_ϵ. We say that a collection $\{v_n\}$ of members of $\mathcal{H}$ indexed by integers *n converges in* (*quadratic*) *norm* to a member v if there exists a function $v \in \mathcal{H}$ such that $\lim_{n\to\infty} \|v - v_n\| = 0$. The additional property which we require for a linear space to be a Hilbert space is that each Cauchy sequence of its elements converges to an element of the space. Because of this property it will be sufficient to check on Cauchy convergence of a sum to assure convergence. In summary:

We say that a set $\mathcal{H}$ is a Hilbert space when:

1. $\mathcal{H}$ is a linear space.
2. An inner product $(\cdot, \cdot)$ is defined on it.
3. Every Cauchy sequence $\{v_n\}$ has a limit in norm (*completeness*).

It turns out that the set of piecewise continuous functions on a finite interval $[0, T]$ is not a Hilbert space, because it is not complete. We call it a *pre-Hilbert space* because it can be completed to a larger space which contains many noncontinuous functions, which are good limits in norm of Cauchy sequences of continuous functions. It is known that this particular $\mathcal{H}$ consists of what mathematicians call "Lebesgue integrable functions" of the interval $[0, T]$, with finite square norm; see a common textbook on this topic for further details (5). One may wonder whether convergence in norm actually entails "pointwise" convergence for each value of t. The Lebesgue theory alluded to in the footnote gives the answer: There is pointwise convergence *almost everywhere*—that is, everywhere except maybe on an unimportant "thin set" of *Lebesgue measure zero* (we say: there is convergence a.e.).

Another nice example of a Hilbert space is the space of quadratically summable simple or doubly infinite series. Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of natural numbers and let $\mathbb{Z} = \{\ldots, -1, 0, 1, 2, \ldots\}$ be the set of integers; then we define $\ell_2^{\mathbb{N}}$ to be the set of (complex valued) sequences $\{a_k : k \in \mathbb{N}$ and $\sum_{k=0}^{\infty} |a_k|^2 < \infty\}$, and similarly for $\ell_2^{\mathbb{Z}}$ (we drop the qualification "$\mathbb{N}$" or "$\mathbb{Z}$" if it either does not matter or it is clear from the context). On ℓ_2 the natural inner product is defined as

$$(a, b) = \sum \overline{b}_k a_k \tag{7}$$

Hilbert spaces are very nice spaces in many respects. They form the natural generalization of the classical finite-dimensional "Euclidean" spaces with their natural quadratic norm (the *Euclidean* norm). In particular, classical features of Euclidean geometry such as orthonormal bases, projection on subspaces, orthogonal decompositions, and so on, generalize. One must be extra cautious when infinite dimensions are involved. Not all infinite-dimensional subspaces of a Hilbert space are Hilbert spaces themselves, because they might not be complete or closed. The orthogonal complement of a subspace is closed, and its orthogonal complement in turn will contain the original space and be a closure of it. Finite-dimensional subspaces, however, are always closed, and luckily they are the ones that occur most frequently in applications.

Of great importance in harmonic analysis are maps between Hilbert spaces. Often the physical situation is such that we are faced with a linear system whose inputs are energy constrained and so are its outputs. The system maps inputs to outputs, often in such a way that the inputed energy is not magnified beyond a certain factor. We call the lower (uniform) limit of such a factor the *norm of the map.* In a later section we shall study it in more detail.

THE DFT, THE FFT, AND CONVOLUTION

The DFT

Given an n-dimensional vector u with components u_k, $k = 0, \ldots, n - 1$ and the complex number $\omega = e^{2\pi j/n}$, we define the

Fourier transform of u (or its harmonic coefficients) as the new vector U whose components are given by

$$U_\ell = \sum_{k=0}^{n-1} \omega^{-\ell k} u_k \tag{8}$$

The Fourier transform can be interpreted as a matrix–vector multiplication given by the expression

$$\begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ \vdots \\ U_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega^{-1} & \omega^{-2} & \cdots & \omega^{-n+1} \\ 1 & \omega^{-2} & \omega^{-4} & \cdots & \omega^{-2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{-n+1} & \omega^{-2(n-1)} & \cdots & \omega^{-(n-1)^2} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{bmatrix} \tag{9}$$

(note that the bottom right entry actually equals ω^{-1}, since $\omega^n = 1$). The matrix which appears in Eq. (9) has special properties: It is a matrix whose rows are mutually orthogonal and so are its columns. Take, for instance, columns r and s; call them f_r and f_s. We have

$$\begin{aligned}(f_r, f_s) &= \textstyle\sum_{\ell=0}^{n-1} \overline{\omega}^{-ls}\omega^{-lr} = \sum_{\ell=0}^{n-1} \omega^{\ell(s-r)} \\ &= \begin{cases} (\text{if } r \neq s) \dfrac{1-\omega^{n(s-r)}}{1-\omega^{(s-r)}} = 0 \\ (\text{if } r = s)\ n \end{cases}\end{aligned} \tag{10}$$

since (1) $\bar{\omega} = \omega^{-1}$ and (2) $\omega^n = 1$.

Let's call F the matrix in the middle of Eq. (9). We have

$$U = Fu \tag{11}$$

If we denote the *hermitian conjugate* of a matrix F by $\tilde{F}$ as the matrix for which $[\tilde{F}]_{ij} = \bar{F}_{ji}$, then the orthogonality says that

$$\tilde{F}F = F\tilde{F} = n.I. \tag{12}$$

As a consequence, the inverse F^{-1} can easily be recovered from F:

$$F^{-1} = \frac{1}{n}\tilde{F} \tag{13}$$

or it can be written out as

$$F^{-1} = \frac{1}{n}\begin{bmatrix} 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega \end{bmatrix} \tag{14}$$

We see that the inverse of a Fourier transform is again a Fourier transform. Each column vector g_k of nF^{-1} is again orthogonal on the others, and we recover u by the formula

$$u = F^{-1}U = \frac{1}{n}\sum_{k=0}^{n-1} g_k U_k = \frac{1}{n}\sum_{k=0}^{n-1}\omega^{kl}U_k \tag{15}$$

The coefficients U_k find an easy interpretation as components for u in the basis formed by the columns of nF^{-1}:

$$(u, g_\ell) = U_\ell \tag{16}$$

Conversely, from Eq. (9) we find

$$U = \sum_{k=0}^{n-1} f_k u_k \tag{17}$$

and hence

$$(U, f_\ell) = n.u_\ell \tag{18}$$

[the discrete Fourier transform (DFT) as commonly defined is not quite symmetrical].

In addition, we obtain an interesting "energy" relation:

$$\|u\|_2^2 = (u,u) = \tilde{u}u = \tilde{U}\tilde{F}FU = \frac{1}{n}\tilde{U}U = \frac{1}{n}(U,U) = \frac{1}{n}\|U\|_2^2 \tag{19}$$

and we see that the quadratic norm of u equals the quadratic norm of U, except for a normalizing factor $1/n$.

The DFT is an orthogonal expansion of a discrete time function with n regularly spaced samples, in an interesting set of basis vectors, the first one giving the constant (DC) component; the second and other basis functions of low index are "slowly varying," while basis functions of high index represent fast variations. Real and imaginary parts of a few of them are given in Fig. 1.

The FFT

The special form of the Fourier transformation matrix leads to additional numerical efficiencies, when executed in the right order. This has been called with the general term *fast Fourier transform* (FFT). We give a quick treatment of the basic Cooley–Tuckey FFT algorithm, and we refer the reader for further discussion of many improvements on the basic algorithm plus connections to Algebra to the excellent treatment given by Blahut (6).

The idea of the FFT is to get as much mileage as possible from the fact that $\omega^n = 1$. Assume that n can be factored into two integers $n = n'n''$. Then, counting indices with small and large steps, we may represent arbitrary indices k, $\ell : 0 \le k$, $\ell \le n-1$ with two small indices (a coarse index and a "vernier" index) using either an n' or n'' "counter" as follows:

$$\begin{aligned} k &= k'_1 + k'_2 n' \\ \ell &= \ell''_1 + \ell''_2 n'' \end{aligned} \tag{20}$$

and identify

$$\begin{aligned} u_{k'_1,k'_2} &= u_k \\ U_{\ell''_1,\ell''_2} &= U_l \end{aligned} \tag{21}$$

Reinterpreting Eq. (8) we find

$$\begin{aligned} U_\ell = U_{\ell''_1,\ell''_2} &= \textstyle\sum_{k'_2=0}^{n''-1}\sum_{k'_1=0}^{n'-1} \omega^{-(\ell''_1+\ell''_2 n'')(k'_1+k'_2 n')} u_{k'_1,k'_2} \\ &= \textstyle\sum_{k'_2=0}^{n''-1}\sum_{k'_1=0}^{n'-1} \omega^{-\ell''_1 k'_1 - \ell''_2 k'_1 n'' - l''_1 k'_2 n'} u_{k'_1,k'_2} \end{aligned} \tag{22}$$

where we have used $\omega^{-n'n''} = 1$.

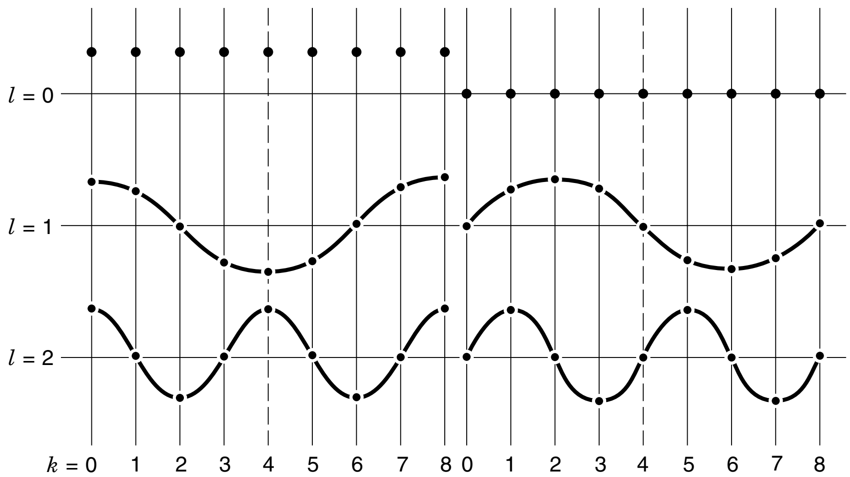

Figure 1. The first few harmonic basis functions for the DFT (case $n = 8$).

Let us also define new coarser "rotors" $\gamma = \omega^{n'}$ and $\beta = \omega^{n''}$, and we find

$$U_\ell = U_{\ell_1'',\ell_2''} = \sum_{k_2'=0}^{n''-1}\sum_{k_1'=0}^{n'-1} \omega^{-\ell_1''k_1'} \cdot \beta^{-\ell_2''k_1'} \cdot \gamma^{-\ell_1''k_2'} u_{k_1',k_2'} \tag{23}$$

If we decide to perform the k_2' summation first, we obtain:

$$U_{\ell_1'',\ell_2''} = \sum_{k_1'=0}^{n'-1} \beta^{-\ell_2''k_1'} \left[\omega^{-\ell_1''k_1'} \sum_{k_2'=0}^{n''-1} \gamma^{-l_1''k_2'} u_{k_1',k_2'} \right] \tag{24}$$

We see that the number of operations is reduced from n^2 (complex) multiplications and additions for the straight DFT, to at most $n(n' + n'' + 1)$ multiplications and $n(n' + n'' - 2)$ additions for the Cooley–Tuckey version. This gain in efficiency happens, however, at the cost of various reshufflings of entries (which in a real-time situation may not be acceptable). The resulting computing scheme for the case $n = 12 = 3 \times 4$ is shown in Fig. 2.

Convolution of Finite Discrete Time Series

In this section we deal with series of length n. If k is an arbitrary integer, then there exists a unique decomposition of k as $k = q \cdot n + r$ in which r and q are integers and $0 \le r \le n - 1$—it is obtained by long division. We say that r is the "residue modulo n" of k, and we write $r = [k]_n$ shorthand. Given two time series $\{u_k\}_{k=0}^{n-1}$ and $\{v_k\}_{k=0}^{n-1}$, we define the following:

The direct convolution (for $i = 0, \ldots, 2n - 1$)

$$w_i = \sum_{k=0}^{i} u_k y_{i-k} \tag{25}$$

The cyclic convolution (for $i = 0, \ldots, n - 1$)

$$w_i = \sum_{k=0}^{n-1} u_k y_{[i-k]_n} \tag{26}$$

One type can be converted to the other by some artifices:

Direct to Cyclic: by zero padding.
Cyclic to Direct: by

$$w_i = \sum_{k=0}^{i} u_k y_{i-k} + \sum_{k=i+1}^{n-1} u_k y_{n+i-k} \tag{27}$$

and observing that both terms are of the direct kind.

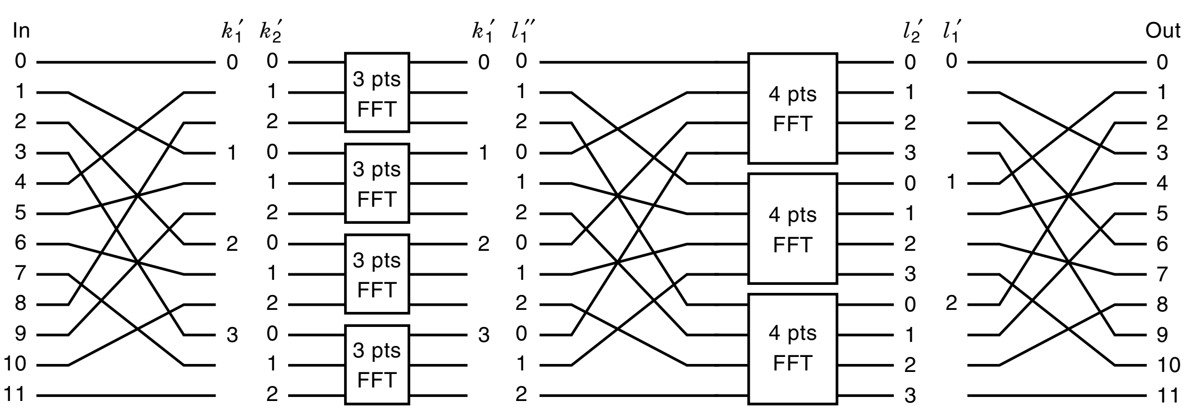

Figure 2. The Cooley–Tuckey fast Fourier transform for $n = 12$.

The *Fourier transform* of a cyclic convolution produces

$$\begin{aligned} W_\ell &= \sum_{i=0}^{n-1} \omega^{-\ell i}\left[\sum_{k=0}^{n-1} u_k y_{[i-k]_n} \right] \\ &= \sum_{k=0}^{n-1} \omega^{-\ell k} u_k \left[\sum_{i=0}^{n-1} \omega^{-\ell [i-k]_n} y_{[i-k]_n} \right] \end{aligned} \tag{28}$$

The last sum is independent of k and evaluates to Y_ℓ. Hence

$$W_\ell = U_\ell Y_\ell \tag{29}$$

We see that the Fourier transform changes convolution to pointwise multiplication. So, if we work systematically in the transform domain, not only the type of operation becomes simpler, but the apparent complexity is also reduced: from n^2 addition and multiplications to just n. For a complete account of complexity, one must of course also add conversion costs; by the FFT those are of the order of $n \log n$.

Multidimensional Discrete Fourier Transforms

Fourier transforms generalize easily to the multidimensional case. We encountered an example when we were treating the FFT. Suppose that $\{u_{ij} : i, j = 0 \ldots n-1\}$ is a two-dimensional array of (possibly complex) values, and let $\omega = e^{2\pi j/n}$ as before; then the two-dimensional Fourier transform is defined as the two-dimensional array:

$$U_{kl} = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \omega^{-(ki+\ell j)} u_{ij} \tag{30}$$

effectively a cascade of Fourier transforms in one dimension (with the other as a parameter). The FFT schemes can of course be applied here as well, and the situation easily generalizes to higher dimensions as well.

THE FOURIER TRANSFORM OF AN ℓ_2 SERIES: CONVOLUTION AND PROPERTIES

The Fourier Transform of an Infinite Series

Suppose now that the length of the time series under consideration increases, $n \to \infty$. Then the "rotor" $\omega = e^{2\pi j/n}$ becomes infinitely small, and the ω^i for $i = 0 \ldots n-1$ represent a very fine sampling of $e^{j\theta}$. For $n = \infty$ it is logical to replace Eq. (8) with the sum

$$U(e^{j\theta}) = \sum_{k=0}^{\infty} e^{-jk\theta} u_k \tag{31}$$

which we shall call the *Fourier transform* $\mathcal{F}(u)$ of u.

Let us first analyze the situation when $u \in \ell_2$—that is, when $\sum_{k=0}^{\infty} |u_k|^2 < \infty$. We show, using the Hilbert space theory we developed in the section entitled "The Basic Framework for Harmonic Analysis: Hilbert Spaces," that finite sums $\sum_{k=0}^{n-1} e^{-jk\theta} u_k$ converge in quadratic norm to a function $U(e^{j\theta})$ which has bounded quadratic norm:

$$\|U(e^{j\theta})\|_2 = \left[\int_{-\pi}^{\pi} |U(e^{j\theta})|^2 \frac{d\theta}{2\pi} \right]^{1/2} \tag{32}$$

(Technically speaking, $U(e^{j\theta})$ turns out to be measurable and square integrable on $\mathbf{T}$. On a set of finite measure such as $\mathbf{T}$ for the Lebesgue measure, functions that belong to $L_2(\mathbf{T})$ are also absolutely integrable, and they belong to $L_1(\mathbf{T})$. This, however, will not be true for functions on sets with infinite measure such as the real line $\mathbb{R}$, on which the spaces L_1 and L_2 both have elements not contained in the other—a major source of difficulty for the continuous time case). Consider the Cauchy sums, for $m < n$:

$$U_{mn} = \sum_{m}^{n-1} e^{-jk\theta} u_k \tag{33}$$

and evaluate their quadratic norm:

$$\begin{aligned} \int_{-\pi}^{\pi} |U_{mn}(e^{j\theta})|^2 \frac{d\theta}{2\pi} &= \sum_{k,l=m}^{n-1} \left[\int_{-\pi}^{\pi} e^{-j(k-\ell)\theta} \frac{d\theta}{2\pi} \right] u_k \overline{u}_\ell \\ &= \sum_{k=m}^{n-1} |u_k|^2 \end{aligned} \tag{34}$$

Since the original series was assumed to be ℓ_2 convergent, we see that also the partial sums U_{0n} of the Fourier transform form a Cauchy series and hence converge to a function $U(e^{j\theta})$, which turns out to be defined pointwise everywhere through this procedure (5). As a bonus, we find conservation of norms or energy. We say that U belongs to the space of square integrable functions (in the sense of Lebesgue) on the unit circle $\mathbf{T}$, $L_2(\mathbf{T})$:

$$\|U\|_2^2 = \int_{\mathbf{T}} |U(e^{j\theta})|^2 \frac{d\theta}{2\pi} = \sum_{k=0}^{\infty} |u_k|^2 = \|u\|_{\ell_2}^2 \tag{35}$$

again a remarkable *Hilbert space isomorphism.* For further information on the analytic properties of L_2 functions and the Fourier transform of infinite series, see Ref. 5.

The definition of the Fourier transform on time series can be extended to the two-sided case. If $\{u_k\}_{-\infty}^{\infty}$ is such a series and if its ℓ_2-norm is finite, we can define in the same way as above

$$U(e^{j\theta}) = \sum_{k=-\infty}^{\infty} e^{-jk\theta} u_k \tag{36}$$

and again we shall have

$$\|U(e^{j\theta})\|_2^2 = \sum_{k=-\infty}^{\infty} |u_k|^2 \tag{37}$$

the preservation of the quadratic norms under the Fourier transform. The coefficients u_k can also easily be recovered by inverse Fourier transform:

$$u_k = \int_{-\pi}^{\pi} e^{jk\theta} U(e^{j\theta}) \frac{d\theta}{2\pi} \tag{38}$$

Convolution of Infinite Time Series

The convolution of two (integrable) functions $U(e^{j\theta})$ and $Y(e^{j\theta})$ on the unit circle $\mathbf{T}$ is given by

$$W(e^{j\theta}) = \int_{-\pi}^{\pi} U(e^{j\xi}) Y(e^{j(\theta-\xi)}) \frac{d\xi}{2\pi} \tag{39}$$

Its Fourier back-transform is luckily easy to compute (the argument parallels the finite-series one above):

$$\begin{aligned} w_k &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{jk\theta}U(e^{j\xi})Y(e^{j(\theta-\xi)})\frac{d\xi}{2\pi}\frac{d\theta}{2\pi} \\ &= \int_{-\pi}^{\pi}\left\{\frac{d\xi}{2\pi}e^{jk\xi}U(e^{j\xi})\cdot\left[\int_{-\pi}^{\pi}\frac{d\theta}{2\pi}e^{jk(\theta-\xi)}Y(e^{j(\theta-\xi)})\right]\right\} \\ &= u_k y_k \end{aligned} \tag{40}$$

All the integrals are finite! The second integral evaluates to y_k because of the circular nature of the integrand, y_k can then be taken out of the external brackets, and the first integral produces u_k.

Hence, convolution in the Fourier domain results in products in the time domain. The converse is true as well, but a little harder to prove because an infinite sum is involved whose convergence is somewhat dubious. We state the result here without proof and refer to the relevant literature for further information (1): If $w_i = \sum_{k=0}^{i} u_k v_{i-k}$, then $W(e^{j\theta}) = U(e^{j\theta})V(ee^{j\theta})$.

Gibbs Phenomenon

The convolution property allows us to study a strange phenomenon typical for quadratic convergence in a Fourier transform domain. Returning to the partial sums $\sum_{k=-n}^{n} e^{-jk\theta}u_k$, we see that their Fourier transforms can be obtained by convolution with the Fourier transform of a block sequence of the type:

$$h^{[n]} = \begin{cases} h_k^{[n]} = 1 & \text{for } |k| \le n \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

Its Fourier transform is

$$\begin{aligned} H^{[n]} &= \textstyle\sum_{-n}^{n} e^{-jk\theta} = \dfrac{e^{-j(n+1)\theta} - e^{jn\theta}}{e^{-j\theta} - 1} \\ &= \dfrac{\sin\left(n + \dfrac{1}{2}\right)\theta}{\sin\dfrac{\theta}{2}} \\ &= \dfrac{\sin\left(n + \dfrac{1}{2}\right)\theta}{\theta/2} \cdot \dfrac{\theta/2}{\sin\theta/2} \end{aligned} \tag{42}$$

and is displayed in Fig. 3 for some large value of n.

The partial sum of a Fourier series equals the Fourier transform of the product of the original series with a block series $h^{[n]}$. Its Fourier transform is thus the convolution of the two transforms. Suppose that the original Fourier transform shows a step for some value θ_0 of θ. The convolution with the transform of the block function will then exhibit the behavior known as the *Gibbs phenomenon* and shown in Fig. 3.

The conclusion is that the partial sum convergence, although pointwise almost everywhere, happens in a rather rough way; and there is actually no convergence in the sense of maximal amplitude deviation, at least not if the functions have discontinuities. The convergence can be improved, however, by using a smoother way of summing. Instead of using a block function to ponder the original series before summation, we could use a triangular function as depicted in Fig. 3. Analytically the transform becomes

$$\sum_{-n}^{n}\left(1 - \frac{|k|}{n}\right)e^{-jk\theta}u_k \tag{43}$$

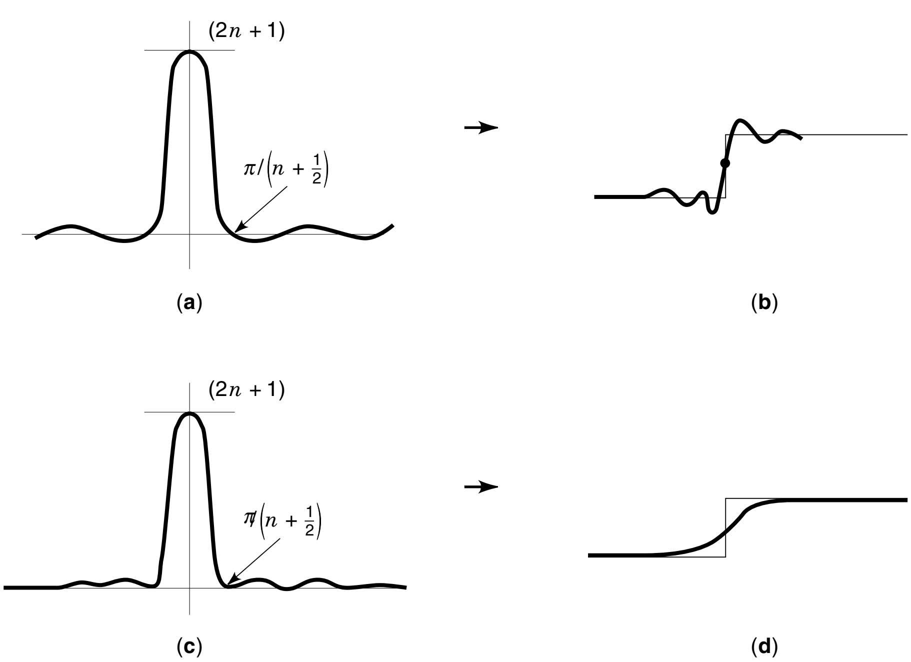

Figure 3. The Gibbs phenomenon: (a) Fourier transform of a block function. (b) The resulting effect on a step in the Fourier transform of the original function. (c) The Fourier transform of a triangular function. (d) The resulting effect on a step.

This sum is called a *Cesaro means.* The Fourier transform of the Cesaro window of order $2n + 1$ can be found, by applying the convolution theorem, as

$$\frac{1}{2n+1}\left(\frac{\sin\left(n+\frac{1}{2}\right)\theta}{\sin\frac{\theta}{2}}\right)^2 \tag{44}$$

and the Gibbs phenomenon has disappeared, since instead of overshooting, the average smooths down the discontinuity, at the cost of a slower rise. The Cesaro window is both a practical summation method and a good tool for convergence proofs.

The *z* Transform

If the time series $\{u_k\}$ is one-sided, then its Fourier transform can still formally be given by Eq. (31), even when convergence is doubtful. Such a one-sided infinite series does make sense purely as a formal series. Introducing the formal variable z^{-1}, we write, independently of convergence,

$$\mathscr{Z}(u) = U(z) = \sum_{k=0}^{\infty} z^{-k} u_k \tag{45}$$

in which the powers of z^{-1}, work as "place holders"—that is, they indicate to which time point the coefficient belongs. $\mathscr{Z}(u)$ is a formal series in z^{-1}. Multiplication of two such formal series yields convolution:

$$\begin{aligned}\mathscr{Z}(u)\mathscr{Z}(v) = {} & u_0v_0 + z^{-1}(u_0v_1 + u_1v_0) \\ & + z^{-2}(u_0v_2 + u_1v_1 + u_2v_0) + \cdots\end{aligned} \tag{46}$$

This is possibly the simplest instance of the linkage between a transform domain and convolution: No convergence or other kind of analytic properties are involved, only the algebraic properties of the *module of formal one-sided power series.*

A further step is to interpret z as a complex variable and study the convergence of the z-transform as a power series in a complex variable. If there are an amplitude A and a rate α such that

$$\forall k\colon\quad |u_k| \le A\alpha^k \tag{47}$$

then we shall have that $U(z)$ is analytic outside the disc $\{z : |z| \le \alpha\}$ in the complex plane $\mathbb{C}$ (it is known from complex analysis that the domain of absolute convergence of a series like Eq. (47) is the complement of a closed disc in $\mathbb{C}$) (we reserve the term *domain* for an open set in the complex plane $\mathbb{C}$).

In case of a *bilateral* series $\{\ldots u_{-2}, u_{-1}, u_0, u_1, \ldots\}$ we can still formally define the series:

$$U(z) = \cdots + z^2u_{-2} + zu_{-1} + u_0 + z^{-1}u_1 + \cdots \tag{48}$$

but then the product of two series may not make any sense anymore, since it will involve infinite summation for the convolution. A further analysis is now necessary to ensure that the product and the convolution of two series do indeed exist.

Decomposing

$$U(z) = U_+(z) + U_-(z) \tag{49}$$

where

$$\begin{aligned}U_+(z) &= u_0 + z^{-1}u_1 + \cdots \\ U_-(z) &= u_{-1} + zu_{-2} + \cdots\end{aligned} \tag{50}$$

we find that $U_+(z)$ will have a domain of absolute convergence of the type $\{z : |z| > \alpha\}$ and $U_-(z)$, one of the type $\{z : |z| < \beta\}$ (they could even be empty!). If $\alpha < \beta$, then there is an annular region $\{z : \alpha < |z| < \beta\}$ in which the double-sided series converges. Finally, if two z-transforms $U(z)$ and $V(z)$ have a common annular convergence domain, then the product $U(z)V(z)$ will be well-defined in that domain.

It is important to note that the domain of convergence and the type of function are intimately related. For example, pending on the domain chosen, the z-transform

$$\frac{1}{(z-\frac{1}{2})(z-2)} \tag{51}$$

may be interpreted as

$$\begin{aligned}&(1)\; z^{-2}\{1 + z^{-1}\tfrac{5}{2} + z^{-2}\tfrac{21}{4} + \cdots\} \\ &(2)\; \cdots - z^2\tfrac{1}{12} - z\tfrac{1}{6} - \tfrac{1}{3} - z^{-1}\tfrac{2}{3} - z^{-2}\tfrac{1}{3} - z^{-3}\tfrac{1}{6} - \cdots \\ &(3)\; \cdots + \tfrac{21}{4}z^2 + \tfrac{5}{2}z + 1\end{aligned} \tag{52}$$

Only one of these interpretations (which are all correct by the way) is properly bounded, namely the one corresponding to the annular region which contains the unit circle of $\mathbb{C}$.

The Hardy Space H_2

Let's return to the one-sided series and assume that it is square integrable, that is, $\sum_{k=0}^{\infty} |u_k|^2 < \infty$. We know already that the Fourier transform $U(e^{j\theta}) = \sum_{k=0}^{\infty} e^{jk\theta}u_k$ exists and is convergent in quadratic norm, as well as pointwise almost everywhere. Its corresponding z-transform $U(z) = \sum_{k=0}^{\infty} z^{-k}u_k$ will be analytic in the region $\{z : |z| > 1\}$—that is, outside the closed unit disc. It turns out that it is characterized by one further property, namely that for $\rho > 1$ the integrals

$$\int_{-\pi}^{\pi} |U(\rho e^{j\theta})|^2 \frac{d\theta}{2\pi} \tag{53}$$

are uniformly bounded by the square of the norms $\|U\|^2_{L_2} = \|u\|^2_{\ell_2}$—they even form a monotonically increasing sequence for $\rho \downarrow 1$.

The space of such Fourier transforms is called the *Hardy space* H_2. It is a (closed) subspace of the space of square integrable functions $L_2(\mathbf{T})$ on the unit circle and plays an important role in many engineering fields such as stochastic system theory and estimation theory. For further information on its properties, see Refs. 5 and 7–9.

The Fourier Transform of a Function on a Finite Interval or of Periodic Functions

A reformulation of the theory of the previous subsections yields a transform theory for functions on a bounded interval, say $[-T/2, T/2]$. One just interchanges time and frequency domains, defining for $f(t) \in L_2([-T/2, T/2])$ the Fourier transform as the series $\{f_k\}_{-\infty}^{\infty}$ with

$$f_k = \int_{-T/2}^{T/2} e^{-jk(2\pi/T)t} f(t) \frac{dt}{T} \tag{54}$$

Reverting to the former theory we identify the interval $[-T/2, T/2]$ with the unit circle in the complex plane. The function $f(t)$ can just as well be assumed to be periodic, and the theory then covers that case as well.

THE FOURIER TRANSFORM OF CONTINUOUS TIME FUNCTIONS

We are now ready to attack the brunt of continuous-time Fourier theory, that is, the Fourier theory for functions of the continuous variable $t \in \mathbb{R}$. Much of the definitions can be motivated by analogy to the previous theory, but now we shall be continuous both in the time domain and in the frequency domain. The theory has some difficulty due to the ill-convergence of some of the integrals used, yet it can be extended to handle very general types of functions (see the section entitled "Distributions or Generalized Functions: A New Framework" for that).

Suppose that $f(t)$ is a (possibly complex) continuous function on $t \in \mathbb{R} = (-\infty, \infty)$, and let's assume that it is absolutely integrable (i.e., $\int_{\mathbb{R}} |f(t)|\, dt < \infty$—we say $f \in L_1(\mathbb{R})$ or that it is bounded in quadratic norm—$f \in L_2(\mathbb{R})$:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty \tag{55}$$

We *define* the *Fourier transform* or *spectrum* of f as a function of the real variable ω on the interval $(-\infty, \infty)$ as

$$\mathscr{F}(f) = F(j\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} f(t)\, dt \tag{56}$$

when $f \in L_1(\mathbb{R})$ or as

$$\mathscr{F}(f) = F(j\omega) = \lim_{T\to\infty} \int_{-T}^{T} e^{-j\omega t} f(t)\, dt \tag{57}$$

when $f \in L_2(\mathbb{R})$. It turns out that the Fourier transform is well-defined, uniformly bounded, and infinitely differentiable in the L_1 case, while it is well-defined a.e. in the L_2 case. A proof of the latter fact parallels the Cauchy convergence proof of the section entitled "The Fourier Transform of an ℓ_2 Series: Convolution and Properties." Details and alternative proofs can be found in classical textbooks (e.g., Refs. 1 and 5). In the L_2 case, $F(j\omega)$ is also square integrable as a function of ω (we'll give a sketch of a proof shortly) and we have norm preservation, known as the *Plancherel identity:*

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(j\omega)|^2 \, d\omega \tag{58}$$

(Note the 2π scaling.)

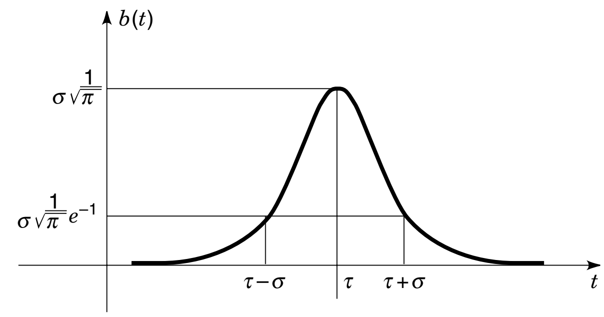

Figure 4. A standard Gaussian or bell-shaped function of unit weight.

Conversely, and concentrating on the L_2 case we find that $f(t)$ can be recovered from $F(j\omega)$, and we have the inversion formula:

$$f(t) = \frac{1}{2\pi} \lim_{\Omega\to\infty} \int_{-j\Omega}^{j\Omega} e^{j\omega t} F(j\omega)\, d\omega \tag{59}$$

(In the L_1 case a stratagem has to be used to recover $f(t)$ from $F(j\omega)$ since the latter may not be integrable. For example, one multiplies with $e^{-(\omega/\sigma)^2}$, takes the Fourier inverse, and lets $\sigma \to \infty$.)

These facts are not so easy to prove. Any book on Fourier transforms has a favorite approach; see, for example, Ref. 1 for a thorough treatment. However, an indirect and quite instructive proof based on the nowadays popular "radial functions" is relatively easy and instructive. We give a sketch of the approach.

Let us start out with the classical bell-shaped or Gaussian functions (also called *radial functions*):

$$b(t) = \frac{1}{\sigma\sqrt{\pi}} e^{-[(t-\tau)/\sigma]^2} \tag{60}$$

These are functions centered at (an arbitrary) real number τ, with unit surface. They have a width approximately equal to σ, and they are infinitely differentiable with derivatives that all go to zero for $t \to \infty$. In the approximation theory of L_2 functions on the real line, one shows that any L_2 function can be approximated as closely as one wishes by a finite sum of weighted (and shifted) bell-shaped functions (they do not form an orthonormal set, but for the argument here that is not necessary). A picture of a radial function is shown in Fig. 4. Because of the fast decay of the bell function, the definition of its Fourier transformation does not necessitate a limiting procedure, and we find reasonably easily (by exponential series expansion or by complex integration) that it is:

$$\mathscr{F}(b) = B(j\omega) = e^{-j\omega\tau} e^{-[\omega/(2/\sigma)]^2} \tag{61}$$

again a bell function, now, however, preceded by a linear phase function characteristic of the time shift in the original bell function.

Checking square surfaces of the functions confirms the norm isomorphism, and the preservation of inner products is also easy to check in this case, because the product of two bell functions is again a bell function. Although two bell functions are not orthogonal on each other, the only important thing is the preservation of inner products, showing equivalence of nonorthogonal bases. Hence we obtain the important *Parseval relation* ('$\mathbb{R}$' is the real line):

$$\int_{t\in\mathbb{R}} \overline{f_2(t)} f_1(t)\,dt = \frac{1}{2\pi}\int_{\omega\in\mathbb{R}} \overline{F_2(j\omega)} F_1(j\omega)\,d\omega \tag{62}$$

again conservation of energy (modulo the correct normalization factor!).

Uncertainty Relation

The width of the bell function considered above was approximately σ, and we see from Eq. (61) that the width of the transform is approximately $2/\sigma$. Hence the product of the widths is a constant 2. This property generalizes to functions concentrated on intervals. A signal that is impulse-like in the time domain will have a wide spectrum, while a signal that is impulse-like in the spectral domain will extend very much in time, such as sine-wave-like functions or harmonics.

Convolution

Suppose that $f_1(t)$ and $f_2(t)$ are two time-domain functions which are absolutely integrable [they are in $L_1(\mathbb{R})$]. Function theory (5) then shows that their convolution integral

$$(f_1 \star f_2)(t) = \int_{\mathbb{R}} f_1(\tau) f_2(t-\tau)\,dt \tag{63}$$

exists and is actually an $L_1(\mathbb{R})$ function itself. Taking Fourier transforms, we find (in the present L_1 case, a direct integration pulls the trick)

$$\mathscr{F}(f_1 \star f_2) = F_1(j\omega)F_2(j\omega) \tag{64}$$

and the Fourier transform of a convolution is again the product of the Fourier transforms of the individual functions.

Conversely, convolution in the spectral domain translates to products in the time domain as follows:

$$\mathscr{F}[f_1(t)\cdot f_2(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} F_1(j\xi)F_2(j(\omega-\xi))\,d\xi \tag{65}$$

(note the factor 2π needed because of the original choice of normalization). Again, proof of these relations has to rely on approximation theory—for example, with bell-shaped functions, for which it is actually easy.

We postpone a further account of Fourier transforms until after the treatment of generalized functions in the section entitled "Distributions or Generalized Functions: A New Framework."

THE LAPLACE TRANSFORM

The Fourier transform as discussed up to this point is restricted to functions which are properly decaying. Although this appears to be fairly essential to make the theory work, it is inconvenient for practical use. Mathematicians and engineers alike have sought a way out of the function theoretical quagmire by deriving a new type of transform which is more tolerant as to the kind of functions treated, and which can be used in a more or less automated way: the Laplace transform.

Let us take, for example, the unit step function at $t = 0$:

$$h(t) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{66}$$

We see that its Fourier summation does not converge. However, if we use a slight (real negative) exponential weight of type $e^{-\alpha_0 t}$ in which α_0 is very small, then the Fourier integral does converge easily and equals $1/(j\omega + \alpha_0)$. So let's, for example, consider the (very large) class of (unilateral) functions $f(t)$ with support on the interval $[0, \infty)$ for which there exists an α_0 such that $f(t)e^{-st} \in L_2([0, \infty))$ for $\Re(s) \geq \alpha_0$. Then we have that for such s

$$F(s) = \int_{\mathbb{R}} e^{-st} f(t)\,dt \tag{67}$$

is well-defined (the integrand is then automatically in L_2 and the integral can be restricted to $[0, \infty)$.) We call $\mathscr{L}(f) = F(s)$ the *Laplace transform* of f. $F(s)$ will be analytic in the region $\Re(s) > \alpha_0$, and the reverse Fourier transform can be applied to it on any vertical line in that region to yield (for $\alpha > \alpha_0$)

$$f(t)e^{-\alpha t} = \frac{1}{2\pi}\int_{\alpha-j\infty}^{\alpha+j\infty} e^{j\omega t} F(\alpha + j\omega)\,d\omega \tag{68}$$

which, with $s = \alpha + j\omega$, becomes

$$f(t) = \frac{1}{2\pi j}\int_{\alpha-j\infty}^{\alpha+j\infty} e^{st} F(s)\,ds \tag{69}$$

and can often be interpreted as a closed-circuit integral in the complex plane. A lot of mileage can be obtained from this fact, making Laplace theory an ideal tool for the analysis of even difficult differential equations (see, for example, Ref. 10 for a good engineering text on this matter). We restrict ourselves here to the analysis of simple discrete-time and continuous-time dynamical systems as an illustration of the use of the theory, and we postpone the treatment of the Laplace transform for generalized functions until the final section, entitled "Distributions or Generalized Functions: A New Framework."

Just as in the case of the z-transform, Eq. (67) may define the Laplace transform of a double-sided $f(t)$, provided that there exists a strip $\{s : \alpha_\ell < s < \beta_r\}$ in $\mathbb{C}$ in which the integrand is integrable. Its inverse will then be given by Eq. (69), in which α is chosen such that $\alpha_\ell < \alpha < \beta_r$. However, L_2-Fourier theory is already richer than the double-sided Laplace theory, and there are many double-sided L_2 functions which do not possess a strip in which they are convergent.

DISCRETE- AND CONTINUOUS-TIME DYNAMICAL SYSTEMS AND THEIR TRANSFORM PROPERTIES

Impulse Responses and Transfer Functions

Transform theory exhibits its full potential only when applied not only to signals but also to transfer operators which map

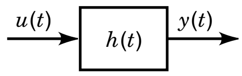

Figure 5. The transfer operator setup.

(input) signals to (output) signals as shown in Fig. 5. We assume that the transfer operator is linear and time-invariant. In that case, the relation between input and output signal can be seen to be a convolution. We start out with the easier discrete time case.

Let us test our system (which we assume initially at rest) with an "impulse"; that is, we apply to it, at a fixed but otherwise arbitrary time i, the impulse (known as the *Kronecker delta*):

$$\delta_i^k = \begin{cases} 1 & \text{for } i = k \\ 0 & \text{otherwise} \end{cases} \tag{70}$$

Because of time invariance, the response will depend only on the time difference $k - i$ and can thus be written h_{k-i}. The response to a one-sided but otherwise arbitrary input applied from time $i = 0$ on will then be given by the convolution

$$y_i = \sum_{k=0}^{i} h_{k-i} u_k \tag{71}$$

because of linearity and time-invariance. Taking z-transforms, we find

$$Y(z) = H(z)U(z) \tag{72}$$

in which $H(z)$ is known as the *transfer function.* It is, by definition, the z-transform of the *impulse response* h_i.

In many cases, a discrete time system can be described by a *state space model,* meaning that there exists a time sequence of vectors $\{x_i\colon i = -\infty, \infty\}$ of some fixed dimension δ internal to the system, along with matrices A, B, C, D which describe the evolution of the system as

$$\begin{aligned} x_{i+1} &= Ax_i + Bu_i \\ y_i &= Cx_i + Du_i \end{aligned} \tag{73}$$

(If u and y are scalar time series, then the dimensions of the matrices A, B, C, D are, respectively, $\delta \times \delta$, $\delta \times 1$, $1 \times \delta$, and 1×1, but vector dimensions for u and y can of course be allowed just as well.)

Using I as a symbol for a unit matrix of appropriate dimensions (from the context), we can z-transform the state equations to obtain an expression for the transfer function in terms of the state space matrices (starting at $t = 0$ with $x_0 = 0$):

$$H(z) = D + C(zI - A)^{-1}B = D + Cz^{-1}(I - z^{-1}A)^{-1}B \tag{74}$$

If all the eigenvalues of A are located inside the closed unit disc, then for $\{z\colon |z| > 1\}$ the series

$$(I - z^{-1}A)^{-1} = I + z^{-1}A + z^{-2}A^2 + \cdots \tag{75}$$

is (absolutely) convergent, in which case one may characterize the system as being *uniformly exponentially stable.* More generally, Eq. (75) can be viewed as the formal one-sided inverse of $(I - z^{-1}A)$, and its rate of increase will be determined by the largest eigenvalue of A (also called its *spectral radius*). In any case, the corresponding impulse response is given by

$$D, CB, CAB, CA^2B, \ldots \tag{76}$$

In the time-continuous case, the situation is more complex. Here also, we wish to write

$$y(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)\,d\tau \tag{77}$$

in which we would like to interpret $h(t)$ as an impulse response. The problem is to give meaning to the notion of *impulse.*

Assume, for the sake of argument, that Eq. (77) is indeed a good representation of the system's behavior seen from the input and output and that $h(t)$ is a reasonably smooth function, uniformly continuous over time. That means that one can find a small interval of size ϵ such that for every t, $h(t)$ is approximately constant in the interval $(t - \epsilon/2, t + \epsilon/2)$. If we now choose an input function $u_\epsilon(t)$, a positive function which is centered around $t = 0$, has support strictly inside an interval $(-\epsilon/2, \epsilon/2)$ of size ϵ, and has unit weight in the sense that

$$\int_{-\epsilon/2}^{\epsilon/2} u_\epsilon(t)\,dt = 1 \tag{78}$$

then the response becomes

$$\begin{aligned} y(t) = \int_{\mathbb{R}} h(t-\tau)u(\tau)\,d\tau &\approx h(t)\int_{-\epsilon/2}^{\epsilon/2} u_\epsilon(\tau)\,d\tau \\ &\approx h(t) \end{aligned} \tag{79}$$

independently of the exact form of u. More general functions can then be represented as weighted sum of impulses, leading, by linearity and time invariance, too Eq. (77).

Conversely, suppose that we have a system for which the response to an impulse-like function as described above is indeed independent of the shape of the impulse, then we can say that the system indeed possesses an *impulse response* $h(t)$. (Think, for example, of a pendulum: When you give it a short hit, its response will depend only on the time integral of the force—the principle of a Coulomb meter) $h(t)$ can of course be derived in other ways—for example, by taking the derivative of the step-response. To cover a larger category of systems, we can be a little more tolerant, and also allow impulselike behavior of the system itself, when you hit it with an impulse, the response consists of a (scaled) impulse, followed by a smooth function h(t). This will, for example, happen to systems that can be described by a state differential equation in the state variable $x(t)$ [we indicate the derivative by $\dot{x}(t)$]:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{80}$$

The analysis of the differential equation shows that the response can be written as

$$y(t) = Du(t) + \int_{-\infty}^{t} Ce^{A(t-\tau)}Bu(\tau)\,d\tau \tag{81}$$

provided that we start with initial state zero and that the integral converges (which will happen if u is smooth enough and has bounded support). Writing $\delta(t)$ for any of the functions $u_\epsilon(t)$ or even their collection, we see that the impulse response then should be written as

$$h(t) = D\delta(t) + Ce^{At}B \tag{82}$$

in which a delta term appears, and that the output to an input $u(t)$ also can be written as a convolution. From Eq. (81) we infer next that the input–output relation of the system can be described by Laplace transforms:

$$Y(s) = H(s)U(s) \tag{83}$$

in which

$$H(s) = D + C(s - A)^{-1}B \tag{84}$$

$\delta(t)$ is known as the *Dirac impulse function;* it is our first example of a generalized function, which we shall explore in more detail in the next section.

Hardy Spaces of Transfer Functions

Returning to the discrete-time domain, let us assume that the transfer map is actually a uniformily bounded input–output map for the quadratic norm. This means that there exists a number M which is larger than or equal to all $\|y\|_2$ when $\|u\|_2 \leq 1$. M is called the *norm of the transfer operator* if it is the smallest of all numbers with that property. Let the system be causal and the output be given by $y_i = \sum_k h_{i-k}u_k$, and let $H(z)$ be the associated transfer function. The interesting question is, To which class does $H(z)$ belong? That is the topic of the Chandrasekharan theorem, which states that $H(z)$ is analytic in the region $\mathbf{E} = \{z : |z| > 1\}$ and is uniformly bounded by M in that region. It is also true that $H(e^{j\theta})$ exists a.e. on the unit circle and is a.e. the radial limit of its values in $\mathbf{E}$. We say that $H(z)$ belongs to the *Hardy space* $H_\infty(\mathbf{E})$, which is a subspace of the space $L_\infty(\mathbf{T})$ of uniformly bounded integrable functions on the unit circle, namely the space of those functions which have a uniformly bounded analytic extension to outside the unit disc.

Even more is true: The uniform norm of $H(e^{j\theta})$ as an $L_\infty(\mathbf{T})$ function is actually equal to its norm as transfer operator:

$$\sup_{\|u\|_2 \leq 1} \|y\|_2 = \sup_{\mathbf{T}} |H(e^{j\theta})| \tag{85}$$

A strong word of caution is appropriate here. Any function in $L_\infty(\mathbf{T})$ is also in $L_2(\mathbf{T})$ and possesses Fourier coefficients with positive and negative indices. Its projection on $H_2(\mathbf{E})$ (i.e., the series obtained by retaining only the Fourier coefficients of non-negative index) is not necessarily an $L_\infty(\mathbf{T})$ function and hence may not belong to $H_\infty(\mathbf{E})$. This is an important fact of Fourier analysis—exemplified, for example, by the ideal lowpass filter, which belongs to L_∞ trivially, but whose real part is not a bounded transfer operator!

In the time-continuous case, much the same facts hold, mutatis mutandis. We also find that a causal input–output transfer map which is uniformly bounded for the quadratic norm on inputs and outputs will result in a transfer function $H(s)$ which is now analytic in the right half-plane $\{s : \Re(s) > 0\}$ and uniformly bounded in that region. In particular, if $H(s)$ is rational, corresponding to a dynamical system as described above, it will have its poles strictly in the open left half-plane.

DISTRIBUTIONS OR GENERALIZED FUNCTIONS: A NEW FRAMEWORK

From the previous treatment of dynamical systems it should be clear that our arsenal of functions is not yet large enough to provide for transforms of generalized functions. To extend it, we need to introduce a new concept, called *distributions.* These are (continuous) functionals—meaning that they achieve their effect when they are used as integrands against other, so-called test functions, which are often chosen as very smooth (for convergence of the integrals). A rich theory of distributions is obtained when the class of test functions are infinitely differentiable functions with compact support (i.e., they are zero outside a finite interval). However, distributions often also make sense when used against less ideal functions.

A formal definition of the Dirac impulse $\delta(t)$ as a distribution can be given as follows: Suppose that $f(t)$ is continuous in a neighborhood of $t = 0$ and that $\mathbf{I}$ is an open interval which contains the point 0, then we say that $\delta(t)$ is the *functional on f which produces the point value* $f(0)$ and we write

$$\int_{\mathbf{I}} \delta(t)f(t)\,dt = f(0) \tag{86}$$

The convention is: we write any functional that makes sense as an integral.

Following this line of thought we can define the derivative of δ on functions $f(t)$ which are differentiable in a neighborhood of $t = 0$ as

$$\int_{\mathbf{I}} \dot{\delta}(t)f(t)\,dt = -\dot{f}(0) \tag{87}$$

(as if we had integrated by parts). And so on for higher-order derivatives.

Fourier and Laplace transforms of such distributions are now easy to define: $\mathscr{F}(\delta) = 1$ and $\mathscr{L}(\delta) = 1$. Also, $\mathscr{F}(\dot{\delta}) = j\omega$ and $\mathscr{L}(\dot{\delta}) = s$.

Another interesting distribution is the function PV$(1/t)$ called *principal value* of $1/t$. On test functions which are continuous and vanish quickly enough at infinity, it produces the functional $\lim_{\epsilon\to 0}[\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}](1/t)f(t)\,dt$.

Many more functionals can be defined. One popular in communication theory is the combfunction, $f(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT)$ for some fixed interval T. A direct definition of the Fourier transform for distributions is in many cases not possible. Following the inventor of distribution theory, Laurent Schwartz (11), we have to resort to a stratagem, using smooth test functions. The test functions that are well-suited for the definition of Fourier transforms are infinitely differentiable and *rapidly decreasing* when their argument goes to infinity, meaning that for all integer powers t^n and all derivatives $f^{(m)}(t)$, $\lim_{|t|\to\infty}$

$t^n f'^{(m)}(t) \to 0$. The Fourier transform $\mathscr{F}(u)$ of such a test function is well-defined by the earlier Fourier theory and one shows that it also belongs to the class of infinitely differentiable and rapidly decreasing functions, this time of ω. If f' is now a functional on such test functions, then the Fourier transform $\mathscr{F}(f')$ can be defined as the distribution which achieves the same functional on test functions as the original distribution does on the Fourier transform of the test functions:

$$\int_{\mathbb{R}} u(\omega)\left[\int_{\mathbb{R}} f'(t)e^{-j\omega t}\,dt\right]d\omega = \int_{\mathbb{R}} f'(t)\left[\int_{\mathbb{R}} u(\omega)e^{-j\omega t}\,d\omega\right]dt \tag{88}$$

Distribution theory gives meaning to convolution as well, and it establishes the convolution theorem for distributions (with some caution, see Ref. 11):

$$\mathscr{F}(u' \star v')(j\omega) = U'(j\omega)V'(j\omega) \tag{89}$$

In addition, distribution theory proves nice convergence properties for functionals: If a sequence of distributions $\{f'_j\}_{j\to\infty}$ has a limit for any test function, then it converges to a distribution f'. These facts give us the necessary tools to handle distributions as objects on which Laplace and Fourier transforms make sense.

As an application, we calculate the Fourier transform of some key distributions and terminate with a list of important Fourier and Laplace transforms.

Let us start with the *Heaviside function* $h(t)$ defined earlier. Its Laplace transform is $1/s$. What is its Fourier transform? One would be tempted to write $1/j\omega$, but that is incorrect. The Fourier transform of $e^{-at}h(t)$ for small a is not suspect:

$$\mathscr{F}(e^{-at}h(t)) = \frac{1}{j\omega + a} = \frac{a}{\omega^2 + a^2} - \frac{j\omega}{\omega^2 + a^2} \tag{90}$$

If we now let $a \to 0$ we see that the first and second terms have no limits in the usual sense (at least not for $\omega = 0$), but they do in the sense of distributions. We find, in that sense,

$$\lim_{a\to 0} \frac{a}{\omega^2 + a^2} = \pi\delta \tag{91}$$

and

$$\lim_{a\to 0} \frac{-j\omega}{\omega^2 + a^2} = PV\left(\frac{1}{j\omega}\right) \tag{92}$$

Figure 6 may help to explain the phenomenon. Hence, the Fourier transform of $h(t)$ is actually $\pi\delta(\omega) + PV(1/j\omega)$.

In a similar vein, the Fourier transform of the constant function $u(t) = 1$ can be found: Using time reversal, we find the Fourier transform for $h(-t)$ as $\pi\delta - PV(1/j\omega)$, and the Fourier transform of 1 becomes $2\pi\delta(\omega)$. For the *signum function* $\text{sgn}(t)$, which equals the sign of t, we find, on the contrary, $2PV(1/j\omega)$. Suppose now that $F(j\omega) = R(j\omega) + jX(j\omega)$ is the decomposition of the Fourier transform of a *real and causal* time function $f(t)$, into real and imaginary parts $R(j\omega)$ and $X(j\omega)$ (it may even be a real distribution). Can we recover X from R and vice versa? The answer is given by realizing that the $R(j\omega)$ is in this case the Fourier transform of the even part of $f(t)$, $f_e(t) = \frac{1}{2}[f(t) + f(-t)]$, while $X(j\omega)$ is the Fourier transform of its odd part, $f_o(t) = \frac{1}{2}[f(t) - f(-t)]$. The even part can be changed to the odd part in the time domain, $f_o(t) = f_e(t)\text{sgn}(t)$, by multiplication with the signum function, which in the Fourier transform domain becomes the convolution:

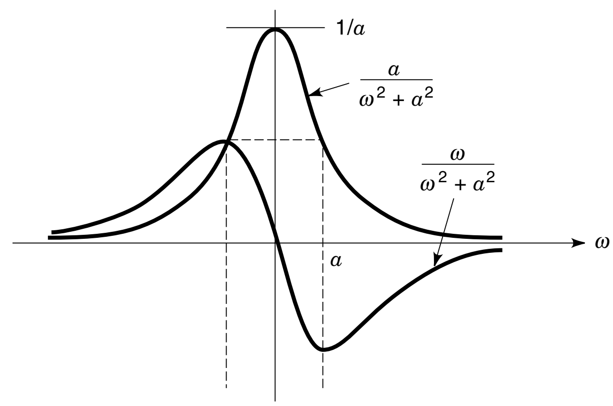

Figure 6. The limiting procedure that leads to the Dirac impulse, and the $PV(1/j\omega)$ distributions.

$$\begin{aligned} X(j\omega) &= -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{R(\xi)}{\omega - \xi}\,d\xi \\ R(j\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{X(\xi)}{\omega - \xi}\,d\xi \end{aligned} \tag{93}$$

These formulas are known as the *Hilbert transform* and have found wide application in communication theory to recover the "quadrature" of a signal which has only been partially modulated (a "single sideband" signal).

BIBLIOGRAPHY

1. H. Dym and H. P. McKean, *Fourier Series and Integrals,* New York: Academic Press, 1972.
2. J. Fourier, *The Analytical Theory of Heat,* translated by A. Freeman, London: Cambridge Univ. Press, 1878. Reprinted by Dover, New York, 1955.
3. W. A. Coppel, J.-B. Fourier—On the occasion of his two hundred birthday, *Amer. Math. Monthly,* **76**: 468–483, 1969.
4. E. W. Hobson, *The Theory of Functions of a Real Variable,* Vol. 2, London: Cambridge Univ. Press, 1926. Reprinted by Dover, 1957.
5. W. Rudin, *Real and Complex Analysis,* New York: McGraw-Hill, 1966.
6. R. E. Blahut, *Fast Algorithms for Digital Signal Processing,* Reading, MA: Addison-Wesley, 1985.
7. K. Hoffman, *Banach Spaces of Analytic Functions,* Englewood Cliffs, NJ: Prentice-Hall, 1962.
8. H. Helson, *Lectures on Invariant Subspaces,* New York: Academic Press, 1964.
9. P. L. Duren, *Theory of* H^p Spaces, New York: Academic Press, 1970.
10. A. Papoulis, *The Fourier Integral and Its Applications,* New York: McGraw-Hill, 1962.
11. L. Schwartz, *Théorie des Distributions,* Paris: Hermann, 1966.

PATRICK DEWILDE
Delft University of Technology

FPGA, CAD FOR. See CAD FOR FIELD PROGRAMMABLE GATE ARRAYS.