Continuous Mathematics

Computer Science Tripos Part IB, Michaelmas Term 1999

J G Daugman

$$\boxed{\int_{-\infty}^{\infty} \Psi_k^*(x)\Psi_j(x)dx = 0}$$

Continuous Mathematics

Computer Science Tripos Part IB, Michaelmas Term
4 lectures by J G Daugman

Aims

The aims of this course are to review some key concepts and operations defined in continuous mathematics involving real- and complex-valued functions of real variables. Focus is on the use and implementation of these notions in the discrete spaces we enter when computing. Topics include: expansions and basis functions; orthogonality and projections; differential equations and their computational solution; linear operators and their eigenfunctions; wavelets and Fourier analysis.

Lectures

- **Review of analysis.** Real and complex-valued functions of a real variable. Power series and transcendental functions. Expansions and basis functions. Smoothness, continuity, limits.
- **Linear vector spaces and decompositions.** Orthogonality, independence, and orthonormality. Linear combinations. Projections, inner products and completeness. Linear subspaces. Useful expansion bases for continuous functions.
- **Differential and integral operators in computation.** The infinitesimal calculus. Taylor series. Numerical integration. Differential equations and computational ways to solve them. Complex exponentials. Introduction to Fourier analysis in one and two dimensions; useful theorems. Convolution and filtering.
- **Signals and systems.** Eigenfunctions of linear operators. Fourier analysis and series; continuous Fourier Transforms and their inverses. Representation in non-orthogonal functions, and wavelets. The degrees-of-freedom in a signal. Sampling theorem. How to operate on continuous signals computationally in order to extract their information.

Objectives

At the end of the course students should:

- Understand how data or functions can be represented in terms of their projections onto other groups of functions.
- Be fluent in the use of, and properties of, complex variables.
- Be able to implement and use, in discrete computational form, such continuous notions as differentiation, integration, and convolution.
- Grasp key properties and uses of Fourier analysis, transforms, and wavelets.

Reference books

Kaplan, W. (1992). *Advanced Calculus.* Addison-Wesley (4th ed.).

Oppenheim, A.V. & Willsky, A.S. (1984). *Signals and Systems.* Prentice-Hall.

1 Purposes of this Course

The discipline of computer science seems to draw mostly upon constructs and operations from discrete mathematics, such as the propositional calculus (logic, syllogisms, truth tables), set membership & relations, and combinatorics. The fundamental notion of an algorithm is a discrete sequence of discrete operations. The elementary hardware devices that implement algorithms are discrete gates, governed by Boolean algebra, and the elementary entities that they manipulate are bits, which are discrete states. Information is generated, transmitted, and stored in discrete form, and everything that happens in computing happens at discrete sequences of points in time – the edges of clock ticks.

So why study continuous mathematics?

Answer 1: *Because the natural world is continuous.* It is the discreteness of digital computing that is unnatural! If we want to model a world that is governed by the laws of physics, we must come to computational terms with continuous processes.

Answer 2: *Because the distinction between discrete and continuous processes is illusory.* Just as every continuous process can be approximated by discrete ones, every discrete process can be modeled as a continuous one.

Answer 3: *Because the two domains are inextricably intertwined, mathematically or physically.* Semiconductor devices such as TTL logical gates really operate through continuous quantities (voltage, current, conductance); continuous theoretical constructs such as differentials and derivatives are only defined in terms of limits of discrete quantities (finite differences); etc.

Answer 4: *Because some of the most interesting and powerful computers that we know about are continuous.* Non-linear dynamical systems in continuous time can be viewed as automata having great computational power; and the most powerful known "computer," the human brain, has the following properties that distinguish it from a digital computer: it lacks numerical calculations; its communications media are stochastic; its components are unreliable and widely distributed; it has no precise connectivity blueprints; and its clocking is asynchronous and extremely slow (milliseconds). Yet its performance in real-time tasks involving perception, learning, and motor control, is unrivaled. As computer scientists we need to be able to study neural processes, and at many levels this requires continuous mathematics.

This short course is intended to be a refresher on some of the major ideas and tools used in continuous mathematics. Its practical purpose within the CST curriculum is to serve as groundwork for the following Pt. II and Diploma courses: *Information Theory and Coding*; *Neural Computing;* and *Computer Vision.*

2 Analysis: Real- and Complex-Valued Functions of a Real Variable

Functions are mappings from some domain to some range. The domain might be the real line (denoted $\mathcal{R}^1$), such as time, or the real plane (denoted $\mathcal{R}^2$), such as an optical image. The range refers to the mapped value or values associated with all the points in the domain. For example, the function might associate to each point on the line or the plane just another real value (a *scalar*, such as temperature), or an ordered set of real values (a *vector*). A weather map showing wind velocity at each point in Britain exemplifies a vector-valued function of the real plane; and so on.

Functions may also associate a complex-valued quantity to each point in the domain. Complex variables are denoted $\mathcal{Z} = a + ib$ where $i = \sqrt{-1}$, and a is the *real part* and b is the *imaginary part* of $\mathcal{Z}$. For example, the Fourier Transform of a musical melody associates a complex variable to every possible frequency, each of which is represented by a point in the (real-valued) frequency domain.

The *complex conjugate* of $\mathcal{Z}$ is denoted by the asterisk (*), and it simply requires changing the sign of the imaginary part. Thus, the complex conjugate of $\mathcal{Z} = a + ib$ is: $\mathcal{Z}^* = a - ib$.

The *modulus* of a complex variable $\mathcal{Z}$ is $\sqrt{a^2 + b^2}$ and it is denoted by $\|\mathcal{Z}\|$. It is easy to see that $\|\mathcal{Z}\| = \sqrt{\mathcal{Z}\mathcal{Z}^*}$.

The *angle* of a complex variable $\mathcal{Z} = a + ib$ is $\tan^{-1}(\frac{b}{a})$ and it is denoted $\angle\mathcal{Z}$. A very important relation that we will use later is: $\mathcal{Z} = \|\mathcal{Z}\| \exp(i\angle\mathcal{Z})$. This can be regarded simply as converting the complex variable $\mathcal{Z}$ from its "Cartesian" form $a + ib$ (where the real part a and the imaginary part b form orthogonal axes defining the complex plane), to polar form (r, θ) in which r is the modulus, or length $\|\mathcal{Z}\|$ of the complex variable, and θ is its angle $\angle\mathcal{Z} = \tan^{-1}(\frac{b}{a})$.

These relations and constructions are central to Fourier analysis and harmonic analysis, which in turn are the mathematical cornerstone of all of electrical engineering involving linear devices; optics; holography; broadcast communications; electronic filter theory; acoustics; quantum mechanics; wave phenomena; much of mechanical engineering, and most of physics! Indeed, the great Nobel Laureate in Physics, Julian Schwinger, once said: "There are only two problems that we can solve in Physics. One is the simple harmonic oscillator [described in terms of the above complex variables]; and the second problem reduces to that one."

3 Power Series and Transcendental Functions

Imagine that you are an 18th-Century astronomer, hard at work on Celestial Mechanics. Understanding and predicting planetary motions requires calculating huge numbers of trigonometric functions such as sine and cosine. Obviously, this is before the age of computers or calculators, or mathematical tables.

How would you compute the sine, the cosine, or the tangent ..., of some angle???

(How *did* they do it??)

Functions such as sine, cosine, logarithm, exponential, hyperbolic cotangent, and so forth, are called *transcendental functions.* They are defined in terms of the limits of *power series*: infinite series of terms involving the argument of the function (the argument of $f(x)$ is x; the argument of $\cos(\theta)$ is θ), raised to an integer power, with associated coefficients in front. Here are some examples of power series that define transcendental functions:

$$\exp(\theta) = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \cdots + \frac{\theta^n}{n!} + \cdots, \tag{1}$$

$$\log(1+\theta) = \theta - \frac{\theta^2}{2} + \frac{\theta^3}{3} - \frac{\theta^4}{4} + \frac{\theta^5}{5} - \cdots, \tag{2}$$

$$\tan(\theta) = \theta + \frac{\theta^3}{3} + 2\frac{\theta^5}{15} + 17\frac{\theta^7}{315} + 62\frac{\theta^9}{2835} + \cdots, \tag{3}$$

$$\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots, \tag{4}$$

$$\sin(\theta) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots, \tag{5}$$

$$\coth(\theta) = \frac{1}{\theta} + \frac{\theta}{3} - \frac{\theta^3}{45} + 2\frac{\theta^5}{945} - \frac{\theta^7}{4725} + \cdots, \tag{6}$$

$$\tag{7}$$

Such expressions – truncated after a certain number of terms – are precisely how computers and calculators evaluate these functions. There is no other way to do it! That is why, if you were the Principal Assistant to the Astronomer Royal in 1720, you spent all of your time with ink quill and paper calculating endless power series such as the above.... :-(

4 Expansions and Basis Functions

The above power series express a function (such as $\sin(x)$ in terms of an infinite series of power functions (like Ax^n) all added together. More generally, almost any function $f(x)$ can be represented perfectly as a linear combination of many other types of functions besides power functions:

$$f(x) = \sum_k a_k \Psi_k(x) \tag{8}$$

where the chosen $\Psi_k(x)$ are called *expansion basis functions.* For example, in the case of Fourier expansions in one dimension, the expansion basis functions are the complex exponentials:

$$\Psi_k(x) = \exp(i\mu_k x) \tag{9}$$

"Finding the representation of some function in a chosen basis" means finding the set of *coefficients* a_k which, when multiplied by their corresponding basis functions $\Psi_k(x)$ and the resulting linear combination of basis functions are summed together, will exactly reproduce the original function $f(x)$.

This is a very powerful tool, because it allows one to choose some universal set of functions in terms of which all other (well-behaved) functions can be represented just as a set of coefficients! In the case of systems analysis, a major benefit of doing this is that knowledge about how members of the chosen universal set of basis functions behave in the system gives one omniscient knowledge about how any possible input function will be treated by the system.

5 Orthogonality, Orthonormality, Inner Products, and Completeness

If the chosen basis functions satisfy the rule that the integral of the conjugate product of any two different members of the family equals zero,

$$\int_{-\infty}^{\infty} \Psi_k^*(x)\Psi_j(x)dx = 0 \quad (k \neq j) \tag{10}$$

then this family of functions is called *orthogonal.*

The above integral is called an *inner product*, and it is often denoted by putting the two functions inside angle brackets (conjugation of one of them is implied:)

$$< \Psi_k(x), \Psi_j(x) > \; \equiv \; \int_{-\infty}^{\infty} \Psi_k^*(x)\Psi_j(x)dx \tag{11}$$

If it is also true that the inner product of any member of this family of functions with itself is equal to 1,

$$< \Psi_k(x), \Psi_j(x) > \; = \; \int_{-\infty}^{\infty} \Psi_k^*(x)\Psi_j(x)dx = 1 \;\; (k = j) \tag{12}$$

then these functions are said to be *orthonormal.* If they form a *complete* basis, then all of the coefficients a_k that are needed to represent some arbitrary function $f(x)$ exactly in terms of the chosen family of orthonormal basis functions $\Psi_k(x)$ can be obtained just by taking the inner products of the original function $f(x)$ with each of the basis functions $\Psi_k(x)$:

$$a_k \;\; = \; < \Psi_k(x), f(x) > \; = \; \int_{-\infty}^{\infty} \Psi_k^*(x)f(x)dx \tag{13}$$

One example of such a representation is the Fourier Transform, which we will examine later.

6 Taylor Series

A particularly powerful and remarkable way to expand a function is simply to use all of its derivatives at some fixed, known, point. It should seem surprising to you that just having complete knowledge about the function at one point, allows you to predict what its value will be at all other points!!

The terms of such an expansion of the function $f(x)$ are based on the successive derivatives of the function at the fixed known point a, denoted $f'(a)$, $f''(a)$, and so forth, each of which is then multiplied by the corresponding power function of the difference between a and the point x at which we desire to know the value of $f(x)$. This is called a Taylor series, and if we consider just the first n terms of such an expansion, then we have *an approximation up to order* n of $f(x)$, which will be denoted $f_n(x)$:

$$f_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{14}$$

7 Continuity and Limits; Derivatives and Anti-Derivatives

The most fundamental notion in continuous mathematics is the idea of a *limit*: the value that an expression inexorably approaches, possibly from below, possibly from above, possibly oscillating around it, tending always closer but possibly never actually reaching it. We have already encountered limits in the power series definitions of transcendental functions. When computers try to calculate something as seemingly straightforward as $\cos(37^o)$, they merely approximate it by *truncating* (considering a finite number of terms in) an infinite series whose *limit* is $\cos(37^o)$. The entire monumental edifice of The Calculus – invented in the decades before 1700 independently by Isaac Newton and Gottfried Leibniz, described by John von Neumann as "the first achievement of modern mathematics, and the greatest technical advance in exact thinking" – is built upon the notion of the limit.

Here are some properties of limits, for continuous functions $f(x)$ and $g(x)$:

$$\lim_{x\to c}[f(x)+g(x)] = \lim_{x\to c}[f(x)] + \lim_{x\to c}[g(x)] \tag{15}$$

$$\lim_{x\to c}[f(x)-g(x)] = \lim_{x\to c}[f(x)] - \lim_{x\to c}[g(x)] \tag{16}$$

$$\lim_{x\to c}[f(x)g(x)] = \lim_{x\to c}[f(x)]\lim_{x\to c}[g(x)] \tag{17}$$

$$\lim_{x\to c}[kf(x)] = k\lim_{x\to c}[f(x)] \tag{18}$$

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \frac{\lim_{x\to c}[f(x)]}{\lim_{x\to c}[g(x)]} \quad \text{assuming } \left(\lim_{x\to c}[g(x)] \neq 0\right) \tag{19}$$

The basic concept of the <u>derivative</u> of a function $f(x)$, denoted $f'(x)$ or $\frac{df(x)}{dx}$, signifying its instantaneous *rate of change* at a point x, is defined as the limit of its <u>Newton Quotient</u> at that point:

$$f'(x) \equiv \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x} \tag{20}$$

The derivative of $f(x)$ exists wherever the above limit exists. It will exist near any point where $f(x)$ is continuous, i.e. if near any point c in the domain of $f(x)$, it is true that $\lim_{x\to c} f(x) = f(c)$.

<u>Review of Rules of Differentiation</u> **(material not lectured)**

- The derivatives of power functions obey a simple rule about exponents:

$$\frac{d}{dx}(x^c) = cx^{c-1} \tag{21}$$

- For any differentiable function $f(x)$ and some constant c,

$$\frac{d}{dx}cf(x) = c\frac{df(x)}{dx} \tag{22}$$

- If u and v are differentiable functions of x, then their sum $u + v$ is a differentiable function of x and

$$\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx} \tag{23}$$

- The product of two differentiable functions u and v is differentiable, and

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx} \tag{24}$$

- If u is some differentiable function of x and c is a constant, then u^c is differentiable, and

$$\frac{d}{dx}(u^c) = cu^{c-1}\frac{du}{dx} \tag{25}$$

- At any point where $v \neq 0$, the quotient u/v of two differentiable functions u and v is itself differentiable, and its derivative is equal to:

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2} \tag{26}$$

- The Chain Rule: if y is a differentiable function of u, and u is a differentiable function of x, then y is a differentiable function of x, and in particular:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} \tag{27}$$

  In another form: if $f(x)$ is differentiable at x, and $g(f(x))$ is differentiable at $f(x)$, then the composite $g \circ f$ is differentiable at x and

$$\frac{d}{dx}g(f(x)) = g'(f(x))f'(x) \tag{28}$$

For a continuous function $f(x)$ that is sampled only at a set of discrete points $\{x_1, x_2, ..., x_n\}$, an estimate of the derivative is called the finite difference. It is defined as you might expect:

$$f'(x) = \frac{f(x_k) - f(x_{k-1})}{(x_k - x_{k-1})} \tag{29}$$

When using a computer to calculate derivatives of continuous data or signals, they must be sampled at a finite number of points; then the above finite difference becomes an estimator of the instantaneous derivative. Clearly, the finite difference approaches the instantaneous derivative in the limit that the sampling interval becomes small: $x_k \to x_{k-1}$.

The area A under a function between two definite points is called its definite integral, and it can be calculated in several ways. Numerically, it can be estimated as the limit of a sum of small rectangular areas inscribed under the function, each of whose height is equal to the value of the function at that point, and whose width Δx shrinks to zero:

$$A = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x \tag{30}$$

Such a summation is the definite integral of the function over the domain covered by the shrinking rectangles, and the origin of the integral sign $\int$ is the letter S in the Latin word Summa, for sum. Thus we denote

$$\int_a^b f(x)dx \equiv \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\Delta x \tag{31}$$

where the set of samples $f(x_k)$ is taken uniformly from $x_1 = a$ to $x_n = b$, and so $\Delta x = (b-a)/n$. The above expression is also termed a Riemann Integral.

Many of the properties we noted earlier for limits obviously apply to definite integrals, since they are themselves defined as limits. For example:

$$\int_a^b kf(x)dx = k\int_a^b f(x)dx \tag{32}$$

$$\int_a^b [f(x)+g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx \tag{33}$$

$$\int_a^b [f(x)-g(x)]dx = \int_a^b f(x)dx - \int_a^b g(x)dx \tag{34}$$

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx \;\; \texttt{if} \;\; f(x) \leq g(x) \;\; \texttt{on} \;\; [a,b] \tag{35}$$

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx \tag{36}$$

The antiderivative of $f(x)$ is denoted $F(x)$ and it is the function whose derivative is $f(x)$, i.e. that function which satisfies

$$\frac{dF(x)}{dx} = f(x) \tag{37}$$

Often one can find the antiderivative of $f(x)$ simply by applying the rules for differentiation in reverse. For example, since we know that if n is a positive integer

$$\frac{d}{dx}(x^n) = nx^{n-1} \tag{38}$$

we can infer that if $f(x) = x^n$, then its antiderivative is:

$$F(x) = \frac{1}{n+1} x^{n+1} \tag{39}$$

Because these are relatively simple symbol-manipulating rules, they can easily be programmed into symbolic math packages such as Stephen Wolfram's famous **Mathematica**, and also **Macsym**, to generate the antiderivatives of even very complicated expressions.

Remarkably, the First Fundamental Theorem of Integral Calculus asserts that in order to calculate the integral of a function $f(x)$ between two points a and b, we need only evaluate its antiderivative $F(x)$ at those two points, and subtract them!

$$\int_a^b f(x)dx = F(b) - F(a) \tag{40}$$

8 Differential Equations and Computational Ways to Solve Them

A vast variety of phenomena that one may wish to model are described in terms of differential equations: algebraic relationships among variables and various orders of their derivatives. The goal is to find the function which satisfies a given differential equation: that function for which the stated relationship among its derivatives etc. is true. Such a function is called a *solution* to the differential equation. For example, the first-order differential equation

$$\frac{d}{dx} f(x) = -\alpha f(x) \tag{41}$$

has the general solution

$$f(x) = A \exp(-\alpha x) \tag{42}$$

(where α may be complex). The second-order differential equation

$$\frac{d^2}{dx^2} f(x) = -\alpha f(x) \tag{43}$$

has solutions such as

$$f(x) = A \cos(\sqrt{\alpha} x), \tag{44}$$

or

$$f(x) = B \sin(\sqrt{\alpha} x), \tag{45}$$

or the more general combination of these sorts of solutions, the complex exponential:

$$f(x) = C \exp(i\sqrt{\alpha} x - i\phi), \tag{46}$$

where we may note that

$$\exp(i\sqrt{\alpha}x - i\phi) = \cos(\sqrt{\alpha}x - \phi) + i\sin(\sqrt{\alpha}x - \phi) \tag{47}$$

Often the solution to a differential equation depends upon initial conditions, or boundary conditions. Sometimes an exact analytic solution can be found, but more generally there is no simple expression for the solution in terms of familiar functions. Rather, one must *numerically solve* the differential equation by writing a program which integrates it, step by step along its variables beginning with the initial conditions. This is one of the major topics of *Numerical Analysis.*

Solving a differential equation (or a coupled family of differential equations) numerically involves the same operations as computing a definite integral by taking the limit of a sum of small rectangles. (That is called Euler's method.) In this respect, computing numerical solutions to differential equations is essentially an exercise in judicious extrapolation. The performance of an algorithm is gauged by its accuracy and its stability when the true solution is rapidly changing; different approaches are needed for different classes of differential equations. We can do better by using local estimators other than the rectangles that we think about as underlying integration when we pass to the limit of infinitesimals. The key issue here is the trade-off between <u>round-off error</u> (which can propagate nastily), and <u>stepsize</u> (i.e. the width of the rectangles), which is denoted h.

- Euler method: next value equals the previous value, plus some stepsize times the derivative at the previous point:

  $$f(x_{k+1}) \approx f(x_k) + hf'(x_k) \tag{48}$$

  The accumulated error ϵ is proportional to stepsize: $\epsilon \approx \mathcal{O}(h)$.

- Improved Euler Method: next value equals the previous value, plus some stepsize times the *average* of the previous value of the derivative and its estimated next value. Error behaviour: accumulated error is proportional to the square of the stepsize: $\epsilon \approx \mathcal{O}(h^2)$.

- Simpson's Method: Use a linear combination of estimates at three points, in an overlapping 1:4:1 sequence. Error behaviour: $\epsilon \approx \mathcal{O}(h^3)$.

- Runge-Kutta Method: Propagate a solution over an interval by combining the information from several Euler-style steps, and then use this information to match a Taylor series expansion up to some specified order (usually 2nd or 4th order terms). Thus the estimate of the solution to the differential equation at any point is a linear combination of these evaluated differentials. Error behaviour: $\epsilon \approx \mathcal{O}(h^4)$.

- Hamming's Predictor-Corrector Method: Examine the recent behaviour of the solution (up to the present point) to make a prediction about the near-term future behaviour, and then correct this prediction using the differential information. Error behaviour: $\epsilon \approx \mathcal{O}(h^5)$.

Numerical instability is the bogey-man when integrating families of differential equations numerically, especially if they happen to be nonlinear or semi-pathological (local behaviour resembling singularities). If the stepsize is too large, then there is gross quantization error. If the stepsize is made too small, then besides the greater computational cost of having to make many more calculations, numerical instability can result from propagation of truncation errors, and the solution is said to "blow-up" (i.e. become unbounded and fail to represent the true solution).

The relationship between the cumulative error ϵ and the stepsize h varies from linear dependence on h for the Euler method, to the fifth power of h for the predictor-corrector method! This reveals the great advantage of choosing a clever method for numerical integration: reducing the stepsize of integration by half can yield a 32-fold reduction in the cumulative error.

To integrate numerically an entire family of coupled differential equations, cycle iteratively through the family, one increment at a time to produce each new estimate of the solution for each member of the family. These new estimates for the whole family at that point are then used in calculating the next differential increment to the solution for each member, and the cycle repeats in a new iteration. Clearly, the fact that the solution to all the equations is required at one point before any of them can be solved at the next point, implies that such numerical solutions are profoundly *serial* and thus generally not amenable to the exploitation of parallel computing architectures across the evolution of the solution. However, parallelism *can* be exploited across the members of the family of equations, with data sharing about the outcome of each successive solution point for each member of the coupled family.

9 Signals and Systems

Many continuous processes or phenomena that one might wish to study or model take the form of *linear time-invariant input-output systems.* Examples include analog electronic circuits, wave phenomena, electromagnetism, optics, and major classes of physical and mathematical systems. These may be represented by a time-varying input $s(t)$, a characteristic and stable "system function" $h(t)$ describing the properties of the system or medium, and a time-varying output response $r(t)$ from the system:

$$s(t) \longrightarrow \boxed{h(t)} \longrightarrow r(t)$$

The study of such systems is called *linear systems analysis*, and it represents one of the key (and most exhaustively understood) areas of continuous mathematics. The next several sections describe the analysis of such systems.

10 Linear Operators and Their Eigenfunctions

The above system $h(t)$ is linear if it obeys the properties of superposition and proportionality:

- Superposition implies that if $r_1(t)$ is the system's response to any input $s_1(t)$, and if $r_2(t)$ is the system's response to any input $s_2(t)$, then the system's response to a third input $s_3(t) = s_1(t) + s_2(t)$ which is the sum of the earlier two inputs, must simply be the sum of its responses to those two inputs separately: $r_3(t) = r_1(t) + r_2(t)$.

- Proportionality implies that if any input $s(t)$ is changed just by multiplying it by a constant k (which may be complex), then the system's original response $r(t)$ simply gets multiplied by the same (possibly complex) constant: $kr(t)$.

Linear systems are thus always described by some *linear operator* $h(t)$. Examples of such linear operators are:

- Any derivative, or combination of derivatives of any order; any linear differential operator with constant coefficients.

- An integral expression.

- A convolution with some fixed waveform.

- Any combination or concatenation of the above.

The *eigenfunctions* of a system are those inputs which emerge completely unchanged at the output, except for multiplication by a constant (which may be complex). A fundamental property of linear systems as described above is that their eigenfunctions are the *complex exponentials* $\exp(i\mu_k t)$:

$$\exp(i\mu_k t) \longrightarrow \boxed{h(t)} \longrightarrow A\exp(i\mu_k t)$$

That is, the only effect which a linear system $h(t)$ can have on an input which is a complex exponential is to multiply it by a complex constant A when generating a response to it. Obviously, other families of input signals would become quite dramatically changed when operated upon by the sorts of linear operators ennumerated above. So, complex exponentials are a very special and important class of functions. In fact, if one can learn how a linear system $h(t)$ responds to all possible complex exponentials (that is to say, if one can measure the complex constant A associated with every possible frequency μ_k of an input complex exponential), then one has complete knowledge about how the system will respond to <u>any</u> other possible input! This is an extraordinary kind of power.

The process works by representing any possible input as a superposition of complex exponentials, and then applying the superposition principles described earlier in order to calculate the output as another linear combination of those same complex exponentials, since they are eigenfunctions. In order to understand and apply this, we need to develop some of the tools of Fourier Analysis.

11 Fourier Analysis

It has been said that the most remarkable and far-reaching relationship in all of mathematics is the simple Euler Relation,

$$e^{i\pi} + 1 = 0 \tag{49}$$

which contains the five most important mathematical constants, as well as harmonic analysis. This simple equation unifies the four main branches of mathematics: $\{0,1\}$ represent arithmetic, π represents geometry, i represents algebra, and $e = 2.718...$ represents analysis, since one way to define e is to compute the limit of $(1 + \frac{1}{n})^n$ as $n \to \infty$.

Fourier analysis is about the representation of functions (or of data, signals, systems, ...) in terms of such complex exponentials. (Almost) any function $f(x)$ can be represented perfectly as a linear combination of basis functions:

$$f(x) = \sum_k a_k \Psi_k(x) \tag{50}$$

where many possible choices are available for the expansion basis functions $\Psi_k(x)$. In the case of Fourier expansions in one dimension, the basis functions are the complex exponentials:

$$\Psi_k(x) = \exp(i\mu_k x) \tag{51}$$

where the complex constant $i = \sqrt{-1}$. A complex exponential contains both a real part and an imaginary part, both of which are simple (real-valued) harmonic functions:

$$\exp(i\theta) = \cos(\theta) + i\sin(\theta) \tag{52}$$

which you can easily confirm by using the power-series definitions for the transcendental functions exp, cos, and sin:

$$\exp(\theta) = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \cdots + \frac{\theta^n}{n!} + \cdots, \tag{53}$$

$$\cos(\theta) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots, \tag{54}$$

$$\sin(\theta) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots, \tag{55}$$

Fourier Analysis computes the complex coefficients a_k that yield an expansion of some function $f(x)$ in terms of complex exponentials:

$$f(x) = \sum_{k=-n}^{k=n} a_k \exp(i\mu_k x) \tag{56}$$

where the parameter μ_k corresponds to frequency and n specifies the number of terms (which may be finite or infinite) used in the expansion.

Each Fourier coefficient a_k in $f(x)$ is computed as the orthonormal projection of the function $f(x)$ onto one complex exponential $\exp(-i\mu_k x)$ associated with that coefficient:

$$a_k = \frac{1}{T}\int_{-T/2}^{+T/2} f(x)\exp(-i\mu_k x)dx \tag{57}$$

where the integral is taken over one period (T) of the function if it is periodic, or from $-\infty$ to $+\infty$ if it is aperiodic. (An aperiodic function is regarded as a periodic one whose period is ∞). For periodic functions the frequencies μ_k used are just all multiples of the repetition frequency; for aperiodic functions, all frequencies must be used. Note that these computed Fourier coefficients a_k are complex-valued. If the function $f(x)$ is real-valued, then its representation in the Fourier domain has two-fold redundancy. The real-parts of the a_k have even-symmetry: $a_k = a_{-k}$, and their imaginary-parts have odd-symmetry: $a_k = -a_{-k}$. Given this "Hermitian" symmetry, only one half of the Fourier coefficients for a real-valued function $f(x)$ need be computed, to obtain them all.

Useful Theorems of Fourier Analysis

Throughout, we will denote the Fourier representation of our original function $f(x)$ as $F(\mu)$, where μ is frequency. If $f(x)$ is a continuous periodic function, then $F(\mu)$ is a discrete set of frequency components in a *Fourier series.* If $f(x)$

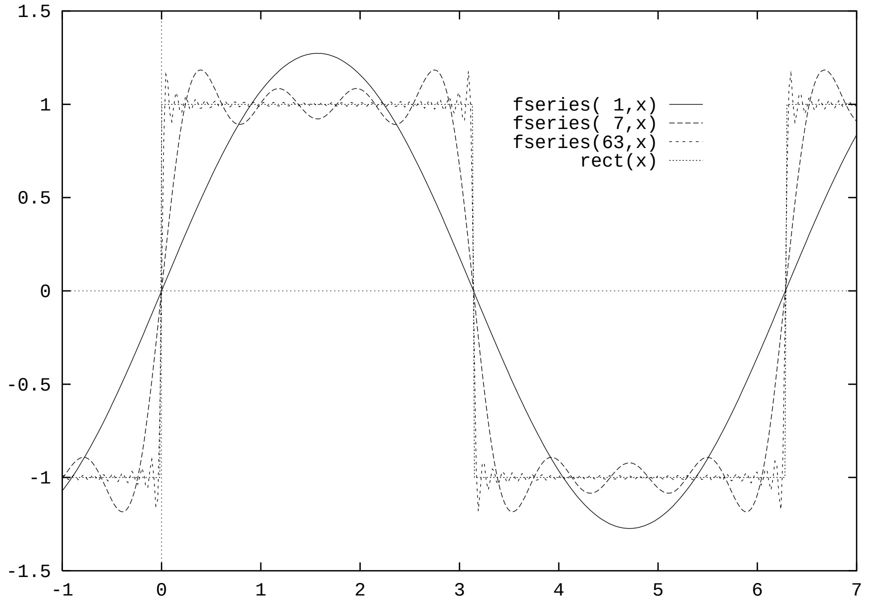

Figure 1: Approximations to rectangular pulse

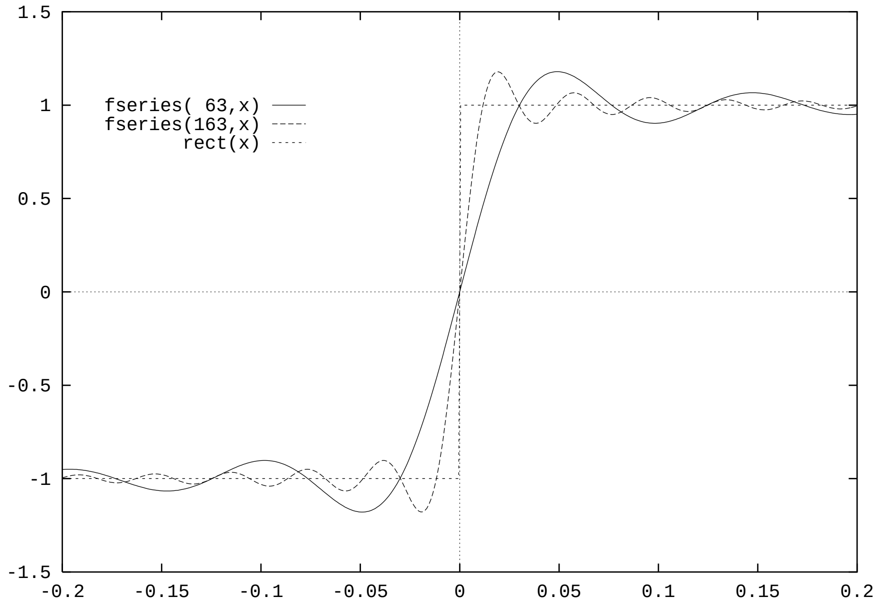

Figure 2: Gibbs phenomenon

Figure 1: Illustration of convergence of Fourier series for a square-wave.

is a continuous but aperiodic function, then $F(\mu)$ is also a continuous and aperiodic function, termed the *Fourier Transform* of $f(x)$.

Shift Theorem: Shifting the original function $f(x)$ by some displacement α merely multiplies its Fourier Transform by $\exp(-i\alpha\mu)$.

Thus the Fourier Transform of the shifted pattern $f(x-\alpha)$ is simply: $F(\mu)\exp(-i\alpha\mu)$.

Similarity Theorem: If the scale of the original function $f(x)$ changes (shrinks or expands along the abscissa) by a factor α, becoming $f(\alpha x)$, then the Fourier Transform of the function also changes (shrinks or expands) by the reciprocal of that factor: $F(\mu) \rightarrow \frac{1}{|\alpha|}F(\mu/\alpha)$.

Convolution Theorem: Let function $f(x)$ have Fourier Transform $F(\mu)$, and let function $g(x)$ have Fourier Transform $G(\mu)$. The convolution of $f(x)$ with $g(x)$, which is denoted $f * g$, combines these two functions to generate a third function $h(x)$, whose value at location (x) is equal to the integral of the product of functions f and g after they undergo a relative shift by amount (x):

$$h(x) = \int_{-\infty}^{+\infty} f(\alpha)g(x-\alpha)d\alpha \tag{58}$$

Thus, convolution is a way of combining two functions, in a sense using each one to blur the other, making all possible relative shifts between the two functions when computing the integral of their product to obtain the corresponding output values.

Convolution is extremely important because it is one basis of describing how any linear system $h(t)$ acts on any input $s(t)$ to generate the corresponding output $r(t)$. Specifically, the output is just the convolution of the input with the characteristic system response function:

$$r(t) = h(t) * s(t) \tag{59}$$

The Convolution Theorem states that convolving any two functions $f(x)$ and $g(x)$ together simply multiplies their two Fourier Transforms together, to generate the Fourier Transform of the result of the convolution:

$$H(\mu) = F(\mu)G(\mu) \tag{60}$$

where $H(\mu)$ is the Fourier Transform of the desired result $h(x)$, and $F(\mu)$ and $G(\mu)$ are the Fourier Transforms of $f(x)$ and $g(x)$, respectively.

This is extremely useful since it is much easier to multiply two functions $F(\mu)$ and $G(\mu)$ together, to obtain $H(\mu)$, than to convolve $f(x)$ and $g(x)$ together to obtain $h(x)$. Of course, exploiting the Convolution Theorem means going into the Fourier Domain and computing the Fourier Transforms of $f(x)$ and $g(x)$, but with powerful and fast FFT algorithms, this is very easy.

<u>Differentiation Theorem</u>: Computing the derivatives of a function $f(x)$ is equivalent to <u>multiplying</u> its Fourier Transform, $F(\mu)$, by frequency raised to a power equal to the <u>order</u> of differentiation:

$$\left(\frac{d}{dx}\right)^m f(x) \stackrel{FT}{\Longrightarrow} (i\mu)^m F(\mu) \tag{61}$$

We will now re-capitulate these theorems for the case of two-dimensional functions $f(x,y)$ because then all these these tools can be applied to computer vision, with $f(x,y)$ being regarded as an image. Many underlying principles and concepts from computer vision (such as scale; edge or motion energy; filtering; directional derivative; textural signature; statistical structure; etc.) must be understood in "spectral" (i.e. Fourier) terms.

In addition to this explanatory role, Fourier analysis can be used directly to construct useful pattern representations that are invariant under translation (change in position), rotation, and dilation (change in size). This is therefore the representation underlying many pattern classification and recognition applications, such as optical character recognition (OCR).

Finally, many operations in practical computing that might not seem related in any way to Fourier analysis, such as computing correlations, convolutions, derivatives, differential equations, and diffusions, are much more easily implemented in the Fourier domain. (Powerful algorithms like the FFT make it easy to go back and forth rapidly between the image and Fourier domains).

Consider a pattern as a distribution over the (x,y) plane: a real-valued two-dimensional function $f(x,y)$.

Any such function can be represented perfectly as a linear combination of two-dimensional basis functions

$$f(x,y) = \sum_k a_k \Psi_k(x,y) \tag{62}$$

where many possible choices are available for the expansion basis functions $\Psi_k(x,y)$. In the case of Fourier expansions in two dimensions, the basis functions are the bivariate complex exponentials:

$$\Psi_k(x,y) = \exp(i(\mu_k x + \nu_k y)) \tag{63}$$

Fourier Analysis computes the coefficients a_k that yield an expansion of the image $f(x,y)$ in terms of bivariate complex exponentials:

$$f(x,y) = \sum_k a_k \exp(i(\mu_k x + \nu_k y)) \tag{64}$$

where the parameters μ_k and ν_k define the coordinates of the 2D Fourier domain. These (μ_k, ν_k) coordinates are called spatial frequency vectors, and the set of them must span the (μ, ν) Fourier plane in a uniform Cartesian lattice.

It is often useful to think of the (μ, ν) Fourier plane as resolved into polar coordinates, where $\omega = \sqrt{\mu^2 + \nu^2}$ is (scalar) <u>spatial frequency</u> and $\phi = \tan^{-1}(\nu/\mu)$ is (scalar) <u>orientation</u>.

Each Fourier coefficient a_k is computed as the orthonormal projection of the entire function $f(x,y)$ onto the vector frequency component $\exp(-i(\mu_k x + \nu_k y))$ associated with that coefficient:

$$a_k = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \exp(-i(\mu_k x + \nu_k y)) dx dy \tag{65}$$

<u>Useful Theorems of Two-Dimensional Fourier Analysis</u>

The following theorems describe what happens to $F(\mu, \nu)$, the 2D Fourier Transform of a function $f(x,y)$, when various operations are applied to $f(x,y)$ before its Fourier Transform is taken.

<u>Shift Theorem</u>: Shifting the original pattern in (x,y) by some 2D displacement (α, β) just multiplies its 2DFT by $\exp(-i(\alpha\mu + \beta\nu))$.

Thus the 2DFT of the shifted pattern $f(x-\alpha, y-\beta)$, is simply: $F(\mu,\nu)\exp(-i(\alpha\mu + \beta\nu))$.

Practical Application: The power spectrum of any pattern is thus <u>translation-invariant</u>: it does not depend on where the pattern is located within the image, and so you don't have to find it first. The power spectrum is defined as the product of the pattern's 2DFT, $F(\mu,\nu)$ times its complex conjugate, $F^*(\mu,\nu)$, which just requires that the sign $(-)$ of the imaginary part of $F(\mu,\nu)$ gets reversed. You can easily see that the power spectrum of the shifted pattern $f(x-\alpha, y-\beta)$, namely:

$$\exp(-i(\alpha\mu+\beta\nu))F(\mu,\nu)\exp(i(\alpha\mu+\beta\nu))F^*(\mu,\nu)$$

is exactly equal to the power spectrum of the original, unshifted pattern: $F(\mu,\nu)F^*(\mu,\nu)$.

Similarity Theorem: If the size of the original pattern $f(x,y)$ changes (shrinks/expands), say by a factor α in the x-direction, and by a factor β in the y-direction, becoming $f(\alpha x, \beta y)$, then the 2DFT of the pattern, $F(\mu,\nu)$, also changes (expands/shrinks) by the reciprocal of those factors: $F(\mu,\nu) \rightarrow \frac{1}{|\alpha\beta|}F(\frac{\mu}{\alpha},\frac{\nu}{\beta})$.

Rotation Theorem: If the original pattern $f(x,y)$ rotates through some angle θ, becoming $f(x\cos(\theta)+y\sin(\theta), -x\sin(\theta)+y\cos(\theta))$, then its 2DFT $F(\mu,\nu)$ also just rotates through the same angle:

$$F(\mu,\nu) \rightarrow F(\mu\cos(\theta)+\nu\sin(\theta), -\mu\sin(\theta)+\nu\cos(\theta)).$$

Practical Application: Size- and orientation-invariant pattern representations can be constructed by these relationships. Specifically, if the Fourier domain (μ,ν) is now mapped into log-polar coordinates (r,θ) where $r = \log(\sqrt{\mu^2+\nu^2})$ and $\theta = \tan^{-1}(\nu/\mu)$, then any dilation (size change) in the original pattern becomes simply a translation along the r-coordinate; and any rotation of the original pattern becomes simply a translation along the orthogonal θ-coordinate in this log-polar Fourier domain. But we saw earlier that translations become immaterial by taking a power spectrum, and so these effects of dilation and rotation of the pattern are eliminated in such a representation.

Combined with the translation-invariant property of the power spectrum, we now see how it becomes possible to represent patterns in a manner that is independent of their position in the image, their orientation, and their size (i.e. the Poincaré group of transformations) These principles are routinely exploited in machine optical character recognition; in military recognition of aircraft profiles; and in "optical computing" generally.

Convolution Theorem: Let function $f(x,y)$ have 2DFT $F(\mu,\nu)$, and let function $g(x,y)$ have 2DFT $G(\mu,\nu)$. The convolution of $f(x,y)$ with $g(x,y)$, which is denoted $f * g$, combines these two functions to generate a third function $h(x,y)$, whose value at location (x,y) is equal to the integral of the product of functions f and g after they undergo a relative shift by amount (x, y):

$$h(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta \tag{66}$$

Thus, convolution is a way of combining two functions, in a sense using each one to blur the other, making all possible relative shifts between the two functions when computing the integral of their product to obtain the corresponding output values.

In the above integral definition, if the minus ($-$) signs were simply replaced with ($+$) signs, the new expression would be the correlation integral.

The Convolution Theorem states that convolving two functions $f(x, y)$ and $g(x, y)$ together in the image domain, simply multiplies their two 2DFT's together in the 2D Fourier domain:

$$H(\mu, \nu) = F(\mu, \nu) G(\mu, \nu) \tag{67}$$

where $H(\mu, \nu)$ is the 2DFT of the desired result $h(x, y)$.

This is extremely useful since it is much easier just to multiply two functions $F(\mu, \nu)$ and $G(\mu, \nu)$ together, to obtain $H(\mu, \nu)$, than to have to convolve $f(x, y)$ and $g(x, y)$ together to obtain $h(x, y)$. Of course, exploiting the Convolution Theorem means going into the 2D Fourier Domain and computing the 2DFT's of $f(x, y)$ and $g(x, y)$, but with powerful and fast 2D-FFT algorithms, this is very easy.

Practical Application: Filtering. The basis of all encoding, image processing, and feature extraction operations is the filtering of an image $f(x, y)$ with some family of filters $g(x, y)$. Filtering is a linear operation implemented by the convolution of an image $f(x, y)$ with filter kernel(s) $g(x, y)$, and the resulting output "image" $h(x, y)$ normally then undergoes non-linear operations of various kinds for image segmentation, motion detection, texture classification, pattern recognition, and image understanding.

Differentiation Theorem: Computing the derivatives of an image $f(x, y)$ is equivalent to multiplying its 2DFT, $F(\mu, \nu)$, by the corresponding frequency coordinate raised to a power equal to the order of differentiation:

$$\left(\frac{d}{dx}\right)^m \left(\frac{d}{dy}\right)^n f(x, y) \stackrel{2DFT}{\Longrightarrow} (i\mu)^m (i\nu)^n F(\mu, \nu) \tag{68}$$

A particularly useful implication of this theorem is that isotropic differentiation, which treats all directions equally (for which the lowest possible order of differentiation is 2nd-order, known as the Laplacian operator ∇^2) is equivalent simply to multiplying the 2DFT of the image by a paraboloid:

$$\nabla^2 f(x,y) \equiv \left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2}\right) f(x,y) \stackrel{2DFT}{\Longrightarrow} -(\mu^2 + \nu^2)F(\mu,\nu) \tag{69}$$

12 The Quantized Degrees-of-Freedom in a Continuous Signal

There are several important results in continuous mathematics expressing the idea that even though a function (such as some time-varying signal) is continuous and dense in time (i.e. the value of the signal is defined at each real-valued moment in time), nevertheless a finite and countable set of discrete numbers suffices to describe it completely, and thus to reconstruct it, provided that its frequency bandwidth is limited.

Such theorems may seem counter-intuitive at first: How could a finite sequence of numbers, at discrete intervals, capture exhaustively the continuous and uncountable stream of numbers that represent all the values taken by a signal over some interval of time?

In general terms, the reason is that bandlimited continuous functions are *not as free* to vary as they might at first seem. Consequently, specifying their values at only certain points, suffices to *determine* their values at all other points.

Some examples are:

- **Nyquist's Sampling Theorem:** If a signal $f(x)$ is strictly bandlimited so that it contains no frequency components higher than W, i.e. its Fourier Transform $F(\mu)$ satisfies the condition

$$F(\mu) = 0 \text{ for } |\mu| > W \tag{70}$$

  then $f(x)$ is completely determined just by sampling its values at a rate of at least $2W$. The signal $f(x)$ can be exactly recovered by using each sampled value to fix the amplitude of a $\text{sinc}(x)$ function,

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \tag{71}$$

  whose width is scaled by the bandwidth parameter W and whose location corresponds to each of the sample points. The continuous signal $f(x)$ can

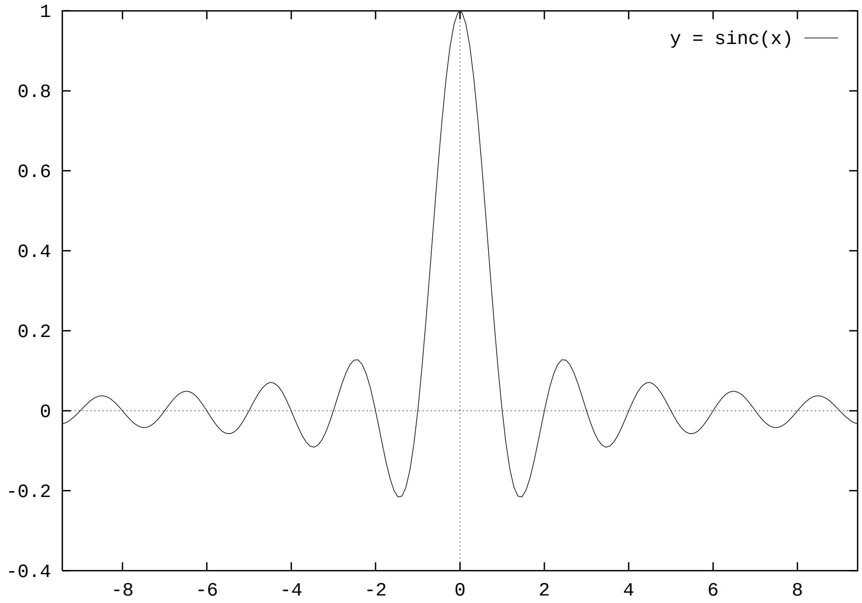

Figure 3: The sinc function, $\frac{sin(\pi x)}{\pi x}$

Figure 2: The sinc function for recovering a continuous signal exactly from its discrete samples, provided their frequency equals the Nyquist rate.

be perfectly recovered from its discrete samples $f_n(\frac{n\pi}{W})$ just by adding all of those displaced sinc(x) functions together, with their amplitudes equal to the samples taken:

$$f(x) = \sum_n f_n\left(\frac{n\pi}{W}\right)\frac{\sin(Wx - n\pi)}{(Wx - n\pi)} \tag{72}$$

(The Figure illustrates this function.) Thus, any signal that is limited in its bandwidth to W, during some duration T has at most $2WT$ degrees-of-freedom. It can be completely specified by just $2WT$ real numbers!

- **The Information Diagram:** The *Similarity Theorem* of Fourier Analysis asserts that if a function becomes narrower in one domain by a factor a, it necessarily becomes broader by the same factor a in the other domain:

$$f(x) \longrightarrow F(\mu) \tag{73}$$

$$f(ax) \longrightarrow |\frac{1}{a}|F(\frac{\mu}{a}) \tag{74}$$

The Hungarian Nobel-Laureate Dennis Gabor took this principle further with great insight and with implications that are still revolutionizing the field of signal processing (based upon wavelets), by noting that an *Information Diagram* representation of signals in a plane defined by the axes of time and frequency is fundamentally *quantized*. There is an irreducible, minimal, volume that any signal can possibly occupy in this plane. Its uncertainty (or spread) in frequency, times its uncertainty (or duration) in time, has an inescapable lower bound.

12.1 Gabor-Heisenberg-Weyl Uncertainty Relation. "Logons."

12.1.1 The Uncertainty Principle

If we define the "effective support" of a function $f(x)$ by its normalized variance, or the normalized second-moment

$$(\Delta x)^2 = \frac{\int_{-\infty}^{+\infty} f(x)f^*(x)(x-x_0)^2 dx}{\int_{-\infty}^{+\infty} f(x)f^*(x)dx} \tag{75}$$

where x_0 is the mean value, or first-moment, of the function

$$x_0 = \frac{\int_{-\infty}^{+\infty} xf(x)f^*(x)dx}{\int_{-\infty}^{+\infty} f(x)f^*(x)dx} \tag{76}$$

and if we similarly define the effective support of the Fourier Transform $F(\mu)$ of the function by its normalized variance in the Fourier domain

$$(\Delta\mu)^2 = \frac{\int_{-\infty}^{+\infty} F(\mu)F^*(\mu)(\mu-\mu_0)^2 d\mu}{\int_{-\infty}^{+\infty} F(\mu)F^*(\mu)d\mu} \tag{77}$$

where μ_0 is the mean value, or first-moment, of the Fourier transform $F(\mu)$

$$\mu_0 = \frac{\int_{-\infty}^{+\infty} \mu F(\mu)F^*(\mu)d\mu}{\int_{-\infty}^{+\infty} F(\mu)F^*(\mu)d\mu} \tag{78}$$

then it can be proven (by Schwartz Inequality arguments) that there exists a fundamental lower bound on the product of these two "spreads," *regardless* of the function $f(x)$!

$$\boxed{(\Delta x)(\Delta\mu) \geq \frac{1}{4\pi}} \tag{79}$$

This is the famous Gabor-Heisenberg-Weyl Uncertainty Principle. Mathematically it is exactly identical to the uncertainty relation in quantum physics, where (Δx) would be interpreted as the position of an electron or other particle, and $(\Delta\mu)$ would be interpreted as its momentum or deBroglie wavelength. We see that this is not just a property of nature, but more abstractly a property of all functions and their Fourier Transforms. It is thus a still further, and more lofty, respect in which the information in continuous signals is quantized, since they must occupy an area in the Information Diagram (time - frequency axes) that is always greater than some irreducible lower bound.

12.1.2 Gabor "Logons"

Dennis Gabor named such minimal areas "logons" from the Greek word for information, or order: logōs. He thus established that the Information Diagram for any continuous signal can contain only a fixed number of information "quanta." Each such quantum constitutes an independent datum, and their total number within a region of the Information Diagram represents the number of independent degrees-of-freedom enjoyed by the signal.

The unique family of signals that actually achieve the lower bound in the Gabor-Heisenberg-Weyl Uncertainty Relation are the complex exponentials multiplied by Gaussians. These are sometimes referred to as "Gabor wavelets:"

$$f(x) = e^{-i\mu_0 x} e^{-(x-x_0)^2/a^2} \tag{80}$$

localized at "epoch" x_0, modulated by frequency μ_0, and with size or spread constant a. It is noteworthy that such wavelets have Fourier Transforms $F(\mu)$ with exactly the same functional form, but with their parameters merely interchanged or inverted:

$$F(\mu) = e^{-ix_0\mu} e^{-(\mu-\mu_0)^2 a^2} \tag{81}$$

Note that in the case of a wavelet (or wave-packet) centered on $x_0 = 0$, its Fourier Transform is simply a Gaussian centered at the modulation frequency μ_0, and whose size is $1/a$, the reciprocal of the wavelet's space constant.

Because of the optimality of such wavelets under the Uncertainty Principle, Gabor (1946) proposed using them as an expansion basis to represent signals. In particular, he wanted them to be used in broadcast telecommunications for encoding continuous-time information. He called them the "elementary functions" for a signal. Unfortunately, because such functions are mutually non-orthogonal, it is very difficult to obtain the actual coefficients needed as weights on the elementary functions in order to expand a given signal in this basis! The first constructive method for finding such "Gabor coefficients" was developed in 1981 by the Dutch physicist Martin Bastiaans, using a dual basis and a complicated non-local infinite series.

When a family of such functions are parameterized to be self-similar, i.e. they are dilates and translates of each other so that they all have a common template ("mother" and "daughter"), then they constitute a (non-orthogonal) *wavelet basis*. Today it is known that an infinite class of wavelets exist which can be used as the expansion basis for signals. Because of the self-similarity property, this amounts to representing or analyzing a signal at different scales. This general field of investigation is called *multi-resolution analysis.*

Two-dimensional Gabor filters over the image domain (x, y) have the functional form

$$f(x, y) = e^{-\left[(x-x_0)^2/\alpha^2+(y-y_0)^2/\beta^2\right]} e^{-i[u_0(x-x_0)+v_0(y-y_0)]} \tag{82}$$

where (x_0, y_0) specify position in the image, (α, β) specify effective width and length, and (u_0, v_0) specify modulation, which has spatial frequency $\omega_0 = \sqrt{u_0^2 + v_0^2}$ and direction $\theta_0 = \arctan(v_0/u_0)$. (A further degree-of-freedom not included above is the relative orientation of the elliptic Gaussian envelope, which creates cross-terms in xy.) The 2-D Fourier transform $F(u, v)$ of a 2-D Gabor filter has exactly the same functional form, with parameters just interchanged or inverted:

$$F(u, v) = e^{-\left[(u-u_0)^2\alpha^2+(v-v_0)^2\beta^2\right]} e^{-i[x_0(u-u_0)+y_0(v-v_0)]} \tag{83}$$

2-D Gabor functions can form a complete self-similar 2-D wavelet expansion basis, with the requirements of orthogonality and strictly compact support relaxed, by appropriate parameterization for dilation, rotation, and translation. If we take $\Psi(x, y)$ to be a chosen generic 2-D Gabor wavelet, then we can generate from this one member a complete self-similar family of 2-D wavelets through the generating function:

$$\Psi_{mpq\theta}(x, y) = 2^{-2m}\Psi(x', y') \tag{84}$$

where the substituted variables (x', y') incorporate dilations in size by 2^{-m}, translations in position (p, q), and rotations through orientation θ:

$$x' = 2^{-m}[x\cos(\theta) + y\sin(\theta)] - p \tag{85}$$

$$y' = 2^{-m}[-x\sin(\theta) + y\cos(\theta)] - q \tag{86}$$

It is noteworthy that as consequences of the similarity theorem, shift theorem, and modulation theorem of 2-D Fourier analysis, together with the rotation isomorphism of the 2-D Fourier transform, all of these effects of the generating function applied to a 2-D Gabor mother wavelet $\Psi(x, y) = f(x, y)$ have corresponding identical or reciprocal effects on its 2-D Fourier transform $F(u, v)$. These properties of self-similarity can be exploited when constructing efficient, compact, multi-scale codes for image structure.

12.1.3 Grand Unification of Domains: an *Entente Cordiale*

Now we can see that the "Gabor domain" of representation actually embraces and unifies both the Fourier domain and the original signal domain! To compute the representation of a signal or of data in the Gabor domain, we find its expansion in terms of elementary functions having the form

$$f(x) = e^{-i\mu_0 x} e^{-(x-x_0)^2/a^2} \tag{87}$$

The single parameter a (the space-constant in the Gaussian term) actually builds a continuous bridge between the two domains: if the parameter a is made very large, then the second exponential above approaches 1.0, and so in the limit our expansion basis becomes

$$\lim_{a\to\infty} f(x) = e^{-i\mu_0 x} \tag{88}$$

the ordinary Fourier basis! If the parameter a is instead made very small, the Gaussian term becomes the approximation to a delta function at location x_o, and so our expansion basis implements pure space-domain sampling:

$$\lim_{\mu_0, a\to 0} f(x) = \delta(x - x_0) \tag{89}$$

Hence the Gabor expansion basis "contains" both domains at once. It allows us to make a continuous deformation that selects a representation lying anywhere on a one-parameter continuum between two domains that were hitherto distinct and mutually unapproachable. A new *Entente Cordiale*, indeed.