Mathematical Methods for Computer Science

6 lectures on Fourier and related methods

Professor J. Daugman

Computer Laboratory
University of Cambridge

Computer Science Tripos, Part IB
Michaelmas Term 2013/14

Outline

- Probability methods (6 lectures, Dr R.J. Gibbens, notes separately)
  - **Limits and inequalities.** (3 lectures)
  - **Markov chains.** (3 lectures)

- Fourier and related methods (6 lectures, Prof. J. Daugman)
  - **Fourier representations.** Inner product spaces and orthonormal systems. Periodic functions and Fourier series. Results and applications. The Fourier transform and its properties. (3 lectures)
  - **Discrete Fourier methods.** The Discrete Fourier transform, efficient algorithms implementing it, and applications. (2 lectures)
  - **Wavelets.** Introduction to wavelets, with applications in signal processing, coding, communications, and computing. (1 lecture)

Reference books

- (*) Pinkus, A. & Zafrany, S.
  Fourier series and integral transforms.
  Cambridge University Press, 1997
- Oppenheim, A.V. & Willsky, A.S.
  Signals and systems.
  Prentice-Hall, 1997

Related on-line video demonstrations:

A tuned mechanical resonator (Tacoma Narrows Bridge): `http://www.youtube.com/watch?v=j-zczJXSxnw`

Interactive demonstrations of convolution: `http://demonstrations.wolfram.com/ConvolutionOfTwoDensities/`

Why Fourier methods are important and ubiquitous

The decomposition of functions (signals, data, patterns, ...) into superpositions of elementary sinusoidal functions underlies much of science and engineering. It allows many problems to be solved.

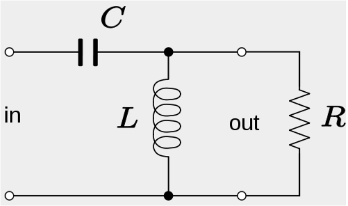
One reason is **Physics:** many physical phenomena such as wave propagation (e.g. sound, water, radio waves) are governed by linear differential operators whose eigenfunctions (unchanged by propagation) are the complex exponentials: $e^{i\omega x} = \cos(\omega x) + i \sin(\omega x)$

Another reason is **Engineering:** the most powerful analytical tools are those of *linear systems analysis*, which allow the behaviour of a linear system in response to *any* input to be predicted by its response to just *certain* inputs, namely those eigenfunctions, the complex exponentials.

A further reason is **Computational Mathematics:** when phenomena, patterns, data or signals are represented in Fourier terms, very powerful manipulations become possible. For example, extracting underlying forces or vibrational modes; the atomic structure revealed by a spectrum; the identity of a pattern under transformations; or the trends and cycles in economic data, asset prices, or medical vital signs.

Simple example of Fourier analysis: analogue filter circuits

Signals (e.g. audio signals expressed as a time-varying voltage) can be regarded as a combination of many frequencies. The relative amplitudes and phases of these frequency components can be manipulated.

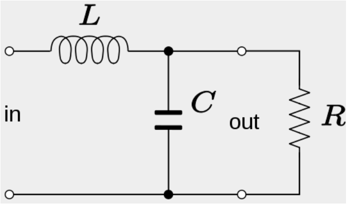


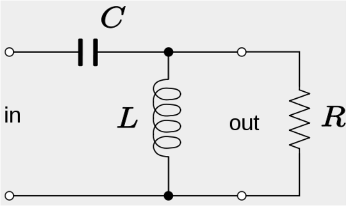
Simple linear analogue circuit elements have a *complex impedance*, Z, which expresses their frequency-dependent behaviour and reveals what sorts of *filters* they will make when combined in various configurations.

- **Resistors** (R in ohms) just have a constant impedance: $Z = R$; but...
- **Capacitors** (C in farads) have low impedance at high frequencies ω, and high impedance at low frequencies: $Z(\omega) = \frac{1}{i\omega C}$
- **Inductors** (L in henrys) have high impedance at high frequencies ω, and low impedance at low frequencies: $Z(\omega) = i\omega L$

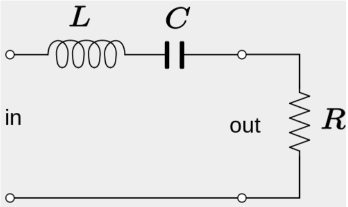
(Simple example of Fourier analysis: filter circuits, con't)

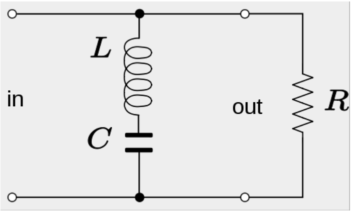
The equations relating voltage to current flow through circuit elements with impedance Z (of which Ohm's Law is a simple example) allow systems to be designed with specific Fourier (frequency-dependent) properties, including filters, resonators, and tuners. Today these would be implemented digitally.





Low-pass filter: higher frequencies are attenuated. High-pass filter: lower frequencies are rejected.





Band-pass filter: only middle frequencies pass. Band-reject filter: middle frequencies attenuate.

So who was Fourier and what was his insight?

Jean Baptiste Joseph Fourier (1768 – 1830)

(Quick biographical sketch of a lucky/unlucky Frenchman)

Orphaned at 8. Attended military school hoping to join the artillery but was refused and sent to a Benedictine school to prepare for Seminary.

The French Revolution interfered. Fourier promoted it, but he was arrested in 1794 because he had then defended victims of the Terror. Fortunately, Robespierre was executed first, and so Fourier was spared.

In 1795 his support for the Revolution was rewarded by a chair at the École Polytechnique. Soon he was arrested again, this time accused of having supported Robespierre. He escaped the guillotine twice more.

Napoleon selected Fourier for his Egyptian campaign and later elevated him to a barony. Fourier was elected to the Académie des Sciences but Louis XVII overturned this because of his connection to Napoleon.

He proposed his famous sine series in a paper on the theory of heat, which was rejected at first by Lagrange, his own doctoral advisor. He proposed the "greenhouse effect." Believing that keeping one's body wrapped in blankets to preserve heat was beneficial, in 1830 Fourier died after tripping in this condition and falling down his stairs. His name is inscribed on the Eiffel Tower.

Mathematical foundations and general framework:

Vector spaces, bases, linear combinations, span, linear independence, inner products, projections, and norms

Inner product spaces

Introduction

In this section we shall consider what it means to represent a function $f(x)$ in terms of other, perhaps simpler, functions.

One example among many is to construct a Fourier series of the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)] \ .$$

How are the coefficients a_n and b_n related to the given function $f(x)$, and how can we determine them?

What other representations might be used?

We shall take a quite general approach to these questions and derive the necessary framework that underpins a wide range of such representations.

Linear space

Definition (Linear space)

A non-empty set V of vectors is a linear space over a field $\mathbb{F}$ of scalars if the following are satisfied.

1. Binary operation $+$ such that if $u, v \in V$ then $u + v \in V$
2. $+$ is associative: for all $u, v, w \in V$ then $(u + v) + w = u + (v + w)$
3. There exists a zero vector, written $\vec{0} \in V$, such that $\vec{0} + v = v$ for all $v \in V$.
4. For all $v \in V$, there exists an inverse vector, written $-v$, such that $v + (-v) = \vec{0}$
5. $+$ is commutative: for all $u, v \in V$ then $u + v = v + u$
6. For all $v \in V$ and $a \in \mathbb{F}$ then $av \in V$ is defined
7. For all $a \in \mathbb{F}$ and $u, v \in V$ then $a(u + v) = au + av$
8. For all $a, b \in \mathbb{F}$ and $v \in V$ then $(a + b)v = av + bv$ and $a(bu) = (ab)u$
9. For all $v \in V$ then $1v = v$, where $1 \in \mathbb{F}$ is the unit scalar.

Choice of scalars

Two common choices of scalar fields, $\mathbb{F}$, are the real numbers, $\mathbb{R}$, and the complex numbers, $\mathbb{C}$, giving rise to real and complex linear spaces, respectively.

The term vector space is a synonym for linear space.

Linear subspace

Definition (Linear subspace)

A subset $W \subset V$ is a linear subspace of V if the W is again a linear space over the same field $\mathbb{F}$ of scalars.

Thus W is a linear subspace if $W \neq \emptyset$ and for all $u, v \in W$ and $a, b \in \mathbb{F}$ we have that $au + bv \in W$.

Linear combinations and spans

Definition (Linear combinations)

If V is a linear space and $v_1, v_2, \ldots, v_n \in V$ are vectors in V then $u \in V$ is a linear combination of $v_1, v_2, \ldots, v_n$ if there exist scalars $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that

$$u = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \, .$$

We also define the span of a set of vectors as all such linear combinations:

$$\mathrm{span}\{v_1, v_2, \ldots, v_n\} = \{u \in V \ : \ u \text{ is a linear combination of } v_1, v_2, \ldots, v_n\} \, .$$

Thus, $W = \mathrm{span}\{v_1, v_2, \ldots, v_n\}$ is a linear subspace of V.

Linear independence

Definition (Linear independence)

Let V be a linear space. The vectors $v_1, v_2, \ldots, v_n \in V$ are linearly independent if whenever

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = \vec{0} \qquad a_1, a_2, \ldots, a_n \in \mathbb{F}$$

then $a_1 = a_2 = \cdots = a_n = 0$

The vectors $v_1, v_2, \ldots, v_n$ are linearly dependent otherwise.

Bases

Definition (Basis)

A finite set of vectors $v_1, v_2, \ldots, v_n \in V$ is a basis for the linear space V if $v_1, v_2, \ldots, v_n$ are linearly independent and $V = \text{span}\{v_1, v_2, \ldots, v_n\}$. The number n is called the dimension of V, written $n = \dim(V)$.

A geometric interpretation and example: any point in the familiar 3 dim Euclidean space $\mathbb{R}^3$ around us can be reached by a linear combination of 3 linearly independent vectors, such as the canonical "(x, y, z) axes." But this would not be possible if the 3 vectors were co-planar; then they would not be linearly independent because any one of them could be represented by a linear combination of the other two, and they would span a space whose dimension is only 2. Note that linear independence of vectors neither requires nor implies orthogonality of the vectors.

A result from linear algebra is that while there are infinitely many choices of basis vectors, any two bases will always consist of the same **number** of element vectors. Thus, the dimension of a linear space is well-defined.

Inner products and inner product spaces

Suppose that V is either a real or complex linear space (that is, the scalars $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$).

Definition (Inner product)

The inner product of two vectors $u, v \in V$, written $\langle u, v \rangle \in \mathbb{F}$, is a scalar value satisfying

1. For each $v \in V$, $\langle v, v \rangle$ is a non-negative real number, so $\langle v, v \rangle \geq 0$
2. For each $v \in V$, $\langle v, v \rangle = 0$ if and only if $v = \vec{0}$
3. For all $u, v, w \in V$ and $a, b \in \mathbb{F}$, $\langle au + bv, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$
4. For all $u, v \in V$ then $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

A linear space together with an inner product is called an inner product space.

Here, $\overline{\langle v, u \rangle}$ denotes the complex conjugate of the complex number $\langle v, u \rangle$. Note that for a real linear space (so, $\mathbb{F} = \mathbb{R}$) the complex conjugate is redundant so the last condition above just says that $\langle u, v \rangle = \overline{\langle v, u \rangle} = \langle v, u \rangle$.

Useful properties of the inner product

Before looking at some examples of inner products there are several consequences of the definition of an inner product that are useful in calculations.

1. For all $v \in V$ and $a \in \mathbb{F}$ then $\langle av, av \rangle = |a|^2 \langle v, v \rangle$
2. For all $v \in V$, $\langle \vec{0}, v \rangle = 0$
3. For all $v \in V$ and finite sequences of vectors $u_1, u_2, \ldots, u_n \in V$ and scalars $a_1, a_2, \ldots, a_n$ then

$$\left\langle \sum_{i=1}^{n} a_i u_i, v \right\rangle = \sum_{i=1}^{n} a_i \langle u_i, v \rangle$$
$$\left\langle v, \sum_{i=1}^{n} a_i u_i \right\rangle = \sum_{i=1}^{n} \overline{a_i} \langle v, u_i \rangle$$

Inner product: examples

Example (Euclidean space, $\mathbb{R}^n$)

$V = \mathbb{R}^n$ with the usual operations of vector addition and multiplication by a real-valued scalar is a linear space over $\mathbb{R}$. Given two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$ we can define an inner product by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i \,.$$

Often this inner product is known as the dot product and is written $x \cdot y$.

Example

Similarly, for $V = \mathbb{C}^n$, we can define an inner product by

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i \overline{y_i} \,.$$

Example (Space of continuous functions on an interval)

$V = C[a, b]$, the space of continuous functions $f : [a, b] \to \mathbb{C}$ with the standard operations of the sum of two functions and multiplication by a scalar, is a linear space over $\mathbb{C}$ and we can define an inner product for $f, g \in C[a, b]$ by

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}dx \,.$$

Note that now the "vectors" have become continuous functions instead. This generalisation can be regarded as the limit in which the number of vector elements becomes infinite, having the density of the reals. The discrete summation over products of corresponding vector elements in our earlier formulation of inner product then becomes, in this limit, a continuous integral of the product of two functions instead.

Norms

The concept of a norm is closely related to an inner product and we shall see that there is a natural way to define a norm given an inner product.

Definition (Norm)

Let V be a real or complex linear space so that, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A norm on V is a function from V to $\mathbb{R}_+$, written $||v||$, that satisfies

1. For all $v \in V$, $||v|| \geq 0$
2. $||v|| = 0$ if and only if $v = \vec{0}$
3. For each $v \in V$ and $a \in \mathbb{F}$, $||av|| = |a| \, ||v||$
4. For all $u, v \in V$, $||u + v|| \leq ||u|| + ||v||$ (the triangle inequality).

A norm can be thought of as a generalisation of the notion of distance, where for any two vectors $u, v \in V$ the number $||u - v||$ is the distance between u and v.

Norms: examples

Example (Euclidean norm)

If $V = \mathbb{R}^n$ or $\mathbb{C}^n$ then for $x = (x_1, x_2, \ldots, x_n) \in V$ define

$$||x|| = +\sqrt{\sum_{i=1}^{n} |x_i|^2}\,.$$

Example (Uniform norm)

If $V = \mathbb{R}^n$ or $\mathbb{C}^n$ then for $x = (x_1, x_2, \ldots, x_n) \in V$ define

$$||x||_\infty = \max\left\{|x_i| \,:\, i = 1, 2, \ldots, n\right\}\,.$$

Example (Uniform norm for continuous functions)

If $V = C[a, b]$ then for each function $f \in V$, define

$$||f||_\infty = \max\left\{|f(x)| \,:\, x \in [a, b]\right\}\,.$$

Cauchy-Schwarz inequality

Theorem (Cauchy-Schwarz inequality)

Let V be a real or complex inner product space with scalars $\mathbb{F}$ then for all $u, v \in V$

$$|\langle u, v\rangle|^2 \leq \langle u, u\rangle \langle v, v\rangle .$$

Proof.

If $v = \vec{0}$ then the result holds trivially. Now assume $v \neq \vec{0}$ so that $\langle v, v\rangle \neq 0$ and let $\lambda \in \mathbb{F}$ then

$$0 \leq \langle u - \lambda v, u - \lambda v\rangle = \langle u, u\rangle - \overline{\lambda}\langle u, v\rangle - \lambda\langle v, u\rangle + |\lambda|^2\langle v, v\rangle$$

Now set $\lambda = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ so the 2nd and 4th terms above cancel, giving

$$0 \leq \langle u, u\rangle - \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle}$$

and hence

$$|\langle u, v\rangle|^2 \leq \langle u, u\rangle\langle v, v\rangle .$$

□

Inner products and norms

Given an inner product space, V, with inner product $\langle \cdot, \cdot \rangle$ there is a natural choice of norm, namely, for all $v \in V$

$$||v|| = +\sqrt{\langle v, v\rangle}\,.$$

Most of the properties that make this a norm follow simply from the properties of the inner product but we shall use the Cauchy-Schwarz inequality to establish the triangle inequality. We have,

$$\begin{aligned} ||u+v||^2 &= \langle u+v, u+v\rangle \\ &= ||u||^2 + \langle u, v\rangle + \langle v, u\rangle + ||v||^2 \\ &\leq ||u||^2 + 2|\langle u, v\rangle| + ||v||^2 \\ &\leq ||u||^2 + 2||u||\,||v|| + ||v||^2 \\ &= (||u|| + ||v||)^2\,. \end{aligned}$$

Hence, the triangle inequality, $||u+v|| \leq ||u|| + ||v||$ holds.

Orthogonal and orthonormal systems

Let V be an inner product space and take the natural choice of norm.

Definition (Orthogonality)

We say that $u, v \in V$ are orthogonal (written $u \perp v$) if $\langle u, v \rangle = 0$.

Definition (Orthogonal system)

A finite or infinite sequence of vectors (u_i) in V is an orthogonal system if

1. $u_i \neq \vec{0}$ for all such vectors u_i
2. $u_i \perp u_j$ for all $i \neq j$.

Definition (Orthonormal system)

An orthogonal system is called an orthonormal system if, in addition, $||u_i|| = 1$ for all such vectors u_i.

A vector $v \in V$ such that $||v|| = 1$ is called a unit vector.

Theorem

Suppose that $\{e_1, e_2, \ldots, e_n\}$ is an orthonormal system in the inner product space V. If $u = \sum_{i=1}^n a_i e_i$ then $a_i = \langle u, e_i \rangle$.

(Another way to say this is that in an orthonormal system, the expansion coefficients are simply the projection coefficients.)

Proof.

$$\begin{aligned} \langle u, e_i \rangle &= \langle a_1 e_1 + a_2 e_2 + \cdots + a_n e_n, e_i \rangle \\ &= a_1 \langle e_1, e_i \rangle + a_2 \langle e_2, e_i \rangle + \cdots + a_n \langle e_n, e_i \rangle \\ &= a_i \, . \end{aligned}$$

$\square$

Hence, if $\{e_1, e_2, \ldots, e_n\}$ is an orthonormal system then for all $u \in \operatorname{span}\{e_1, e_2, \ldots, e_n\}$ we have

$$u = \sum_{i=1}^{n} a_i e_i = \sum_{i=1}^{n} \langle u, e_i \rangle e_i \, .$$

Generalized Fourier coefficients

Let V be an inner product space and $e_1, e_2, \ldots, e_n$ an orthonormal system (n being finite or infinite).

Definition (Generalized Fourier coefficients)

Given a vector $u \in V$, the scalars $\langle u, e_i \rangle$ ($i = 1, 2, \ldots, n$) are called the Generalized Fourier coefficients of u with respect to the given orthonormal system.

These coefficients are generalized in the sense that they refer to a general orthonormal system. It is not assumed that the vectors e_i are actually complex exponentials, the Fourier basis.

Orthogonal projections

Suppose that V is an inner product space and $e_1, e_2, \ldots, e_n$ is an orthonormal system. Define $W = \operatorname{span}\{e_1, e_2, \ldots, e_n\}$ and let $u \in V$ be any vector. We have seen that for $u \in W$

$$u = \sum_{i=1}^{n} \langle u, e_i \rangle e_i$$

but if $u \notin W$ then certainly

$$u \neq \sum_{i=1}^{n} \langle u, e_i \rangle e_i$$

since u is not a linear combination of the vectors $e_1, e_2, \ldots, e_n$. Nevertheless, there is a close connection between u and the expression $\sum_{i=1}^{n} \langle u, e_i \rangle e_i$.

Definition (Orthogonal projection)

For all $u \in V$ we define the orthogonal projection of u in W, $\tilde{u}$, by

$$\tilde{u} = \sum_{i=1}^{n} \langle u, e_i \rangle e_i \, .$$

Infinite orthonormal systems

We now consider the situation of an inner product space, V, with $\dim(V) = \infty$ and consider orthonormal systems $\{e_1, e_2, \ldots\}$ consisting of infinitely many vectors.

Definition (Convergence in norm)

Let $\{u_1, u_2, \ldots\}$ be an infinite sequence of vectors in the normed linear space V and let $\{a_1, a_2, \ldots\}$ be a sequence of scalars. We say that the series

$$\sum_{n=1}^{\infty} a_n u_n$$

converges in norm to $w \in V$ if

$$\lim_{m \to \infty} ||w - \sum_{n=1}^{m} a_n u_n|| = 0 \, .$$

Closure

Consider an infinite orthonormal system $\{e_1, e_2, \dots\}$ in an inner product space V.

Definition (Closed)

The system is called closed in V if for all $u \in V$

$$\lim_{m\to\infty} ||u - \sum_{n=1}^{m} \langle u, e_n \rangle e_n|| = 0\,.$$

Remarks on closure

- If a system is not closed then there must exist some $u \in V$ such that the linear combination
$$\sum_{n=1}^{m} \langle u, e_n \rangle e_n$$
cannot be made arbitrarily close to u, for all choices of m.

(Remarks on closure, con't)

- If the system is closed it may still be that the required number of terms in the above linear combination for a "good" approximation is too great for practical purposes.
- Seeking alternative closed systems of orthonormal vectors may produce "better" approximations in the sense of requiring fewer terms for a given accuracy.
- There exists a numerical method of constructing an orthonormal system $\{e_1, e_2, \ldots\}$ such that any given set of vectors $\{u_1, u_2, \ldots\}$ (which are often a set of multivariate data) can be represented within it with the best possible accuracy using any specified finite **number** of terms. Optimising the approximation under truncation requires deriving the orthogonal system $\{e_1, e_2, \ldots\}$ **from** the data set $\{u_1, u_2, \ldots\}$. This is called the Karhunen-Loève transform or alternatively the Hotelling transform or Dimensionality Reduction or Principal Components Analysis, and it is used in statistics and in exploratory data analysis, but it is outside the scope of this course.

Fourier series

Representing functions

In seeking to represent functions as linear combinations of simpler functions we shall need to consider spaces of functions with closed orthonormal systems.

Definition (piecewise continuous)

A function is piecewise continuous if it is continuous, except at a finite number of points and at each such point of discontinuity, the right and left limits exists and are finite.

The space, E, of piecewise continuous functions $f : [-\pi, \pi] \to \mathbb{C}$ is seen to be a linear space, under the convention that we regard two functions in E as identical if they are equal at all but a finite number of points. We consider the functions over the interval $[-\pi, \pi]$ for convenience.

For $f, g \in E$, then

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx$$

defines an inner product on E.

A closed infinite orthonormal system for E

An important result is that

$$\left\{\frac{1}{\sqrt{2}}, \sin(x), \cos(x), \sin(2x), \cos(2x), \sin(3x), \cos(3x), \ldots\right\}$$

is a closed infinite orthonormal system in the space E.

Here we shall just demonstrate orthonormality and omit establishing that this system is closed.

Writing

$$||f|| = +\sqrt{< f, f >}$$

as the norm associated with our inner product, it can be establish that

$$||\frac{1}{\sqrt{2}}||^2 = 1$$

and similarily that for each $n = 1, 2, \dots$

$$||\sin(nx)||^2 = ||\cos(nx)||^2 = 1$$

and that for $m, n \in \mathbb{N}$

- $\langle \frac{1}{\sqrt{2}}, \sin(nx) \rangle = 0$
- $\langle \frac{1}{\sqrt{2}}, \cos(nx) \rangle = 0$
- $\langle \sin(mx), \cos(nx) \rangle = 0$
- $\langle \sin(mx), \sin(nx) \rangle = 0,\ m \neq n$
- $\langle \cos(mx), \cos(nx) \rangle = 0,\ m \neq n$.

Fourier series

From our knowledge of closed orthonormal systems $\{e_1, e_2, \ldots\}$ we know that we can represent any function $f \in E$ by a linear combination

$$\sum_{n=1}^{\infty} \langle f, e_n \rangle e_n \, .$$

We now turn to consider the individual terms $\langle f, e_n \rangle e_n$ in the case of the closed orthonormal system

$$\left\{ \frac{1}{\sqrt{2}}, \sin(x), \cos(x), \sin(2x), \cos(2x), \sin(3x), \cos(3x), \ldots \right\} .$$

There are three cases, either $e_n = \frac{1}{\sqrt{2}}$ or $\sin(nx)$ or $\cos(nx)$. Recall that the vectors e_n are actually functions
in $E = \{f : [-\pi, \pi] \to \mathbb{C} : f \text{ is piecewise continuous}\}$

If $e_n = 1/\sqrt{2}$ then

$$\langle f, e_n \rangle e_n = \frac{1}{\pi} \left(\int_{-\pi}^{\pi} f(t) \frac{1}{\sqrt{2}} dt \right) \frac{1}{\sqrt{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt \, .$$

If $e_n = \sin(nx)$ then

$$\langle f, e_n \rangle e_n = \frac{1}{\pi} \left(\int_{-\pi}^{\pi} f(t) \sin(nt) \, dt \right) \sin(nx) \, .$$

If $e_n = \cos(nx)$ then

$$\langle f, e_n \rangle e_n = \frac{1}{\pi} \left(\int_{-\pi}^{\pi} f(t) \cos(nt) \, dt \right) \cos(nx) \, .$$

Fourier coefficients

Thus the linear combination

$$\sum_{n=1}^{\infty} \langle f, e_n \rangle e_n$$

becomes the familiar Fourier series for a function f, namely

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right]$$

where

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)\, dx, \quad n = 0, 1, 2, \ldots$$
$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)\, dx, \quad n = 1, 2, 3, \ldots.$$

Note how the constant term is written $a_0/2$ where $a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx$.

Periodic functions

Our Fourier series

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

defines a function, $g(x)$, say, that is 2π-periodic in the sense that

$$g(x + 2\pi) = g(x), \qquad \text{for all } x \in \mathbb{R}\,.$$

Hence, it is convenient to extend $f \in E$ to a 2π-periodic function defined on $\mathbb{R}$ instead of being restricted to $[-\pi, \pi]$.

Even and odd functions

A particularly useful simplification occurs when the function $f \in E$ is either an even function, that is, for all x,

$$f(-x) = f(x)$$

or an odd function, that is, for all x,

$$f(-x) = -f(x)\,.$$

The following properties can be easily verified.

1. If f, g are even then fg is even
2. If f, g are odd then fg is even
3. If f is even and g is odd then fg is odd
4. If g is odd then for any $h > 0$ then $\int_{-h}^{h} g(x)dx = 0$
5. If g is even then for any $h > 0$ then $\int_{-h}^{h} g(x)dx = 2\int_{0}^{h} g(x)dx$.

Even functions and cosine series

Recall that the Fourier coefficients are given by

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)\, dx, \quad n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)\, dx, \quad n = 1, 2, 3, \ldots$$

so if f is even then they become

$$a_n = \frac{2}{\pi} \int_{0}^{\pi} f(x) \cos(nx)\, dx, \quad n = 0, 1, 2, \ldots$$

$$b_n = 0, \quad n = 1, 2, 3, \ldots .$$

Odd functions and sine series

Similarly, the Fourier coefficients

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,dx, \quad n = 0, 1, 2, \ldots$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,dx, \quad n = 1, 2, 3, \ldots,$$

for the case where f is an odd function become

$$a_n = 0, \quad n = 0, 1, 2, \ldots$$
$$b_n = \frac{2}{\pi}\int_{0}^{\pi} f(x)\sin(nx)\,dx, \quad n = 1, 2, 3, \ldots .$$

Fourier series: example 1

Consider $f(x) = x$ for $x \in [-\pi, \pi]$ then f is clearly odd and so we need to calculate a sine series with coefficients, b_n, $n = 1, 2, \ldots$ given by

$$
\begin{aligned}
b_n = \frac{2}{\pi}\int_0^\pi x\sin(nx)\,dx &= \frac{2}{\pi}\left\{\left[-x\frac{\cos(nx)}{n}\right]_0^\pi + \int_0^\pi \frac{\cos(nx)}{n}dx\right\} \\
&= \frac{2}{\pi}\left\{-\pi\frac{(-1)^n}{n} + \left[\frac{\sin(nx)}{n^2}\right]_0^\pi\right\} \\
&= \frac{2}{\pi}\left\{-\pi\frac{(-1)^n}{n} + 0\right\} = \frac{2(-1)^{n+1}}{n}\,.
\end{aligned}
$$

Hence the Fourier series of $f(x) = x$ is

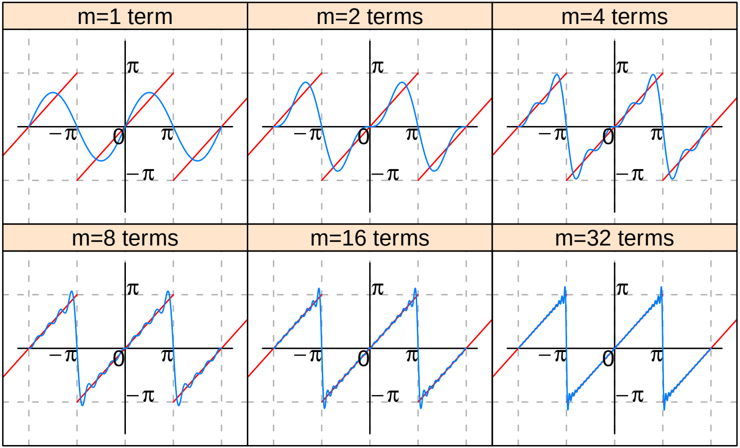
$$
\sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n}\sin(nx)\,.
$$

Observe that the series does not agree with $f(x)$ at $x = \pm\pi$, the endpoints of the interval — a matter that we shall return to later.

(example 1, con't)

Let us examine plots of the partial sums to m terms

$$\sum_{n=1}^{m} \frac{2(-1)^{n+1}}{n} \sin(nx)\,.$$



Fourier series: example 2

Now suppose $f(x) = |x|$ for $x \in [-\pi, \pi]$ which is clearly an even function so we need to construct a cosine series with coefficients

$$a_0 = \frac{2}{\pi}\int_0^{\pi} x dx = \frac{2}{\pi}\frac{\pi^2}{2} = \pi$$

and for $n = 1, 2, \ldots$

$$\begin{aligned} a_n &= \frac{2}{\pi}\int_0^{\pi} x\cos(nx)\,dx = \frac{2}{\pi}\left\{\left[\frac{x\sin(nx)}{n}\right]_0^{\pi} - \int_0^{\pi}\frac{\sin(nx)}{n}dx\right\} \\ &= \frac{2}{\pi}\left\{\left[\frac{\cos(nx)}{n^2}\right]_0^{\pi}\right\} = \frac{2}{\pi}\left\{\frac{(-1)^n - 1}{n^2}\right\} = \begin{cases} -\frac{4}{\pi n^2} & n \text{ is odd} \\ 0 & n \text{ is even} \end{cases} . \end{aligned}$$

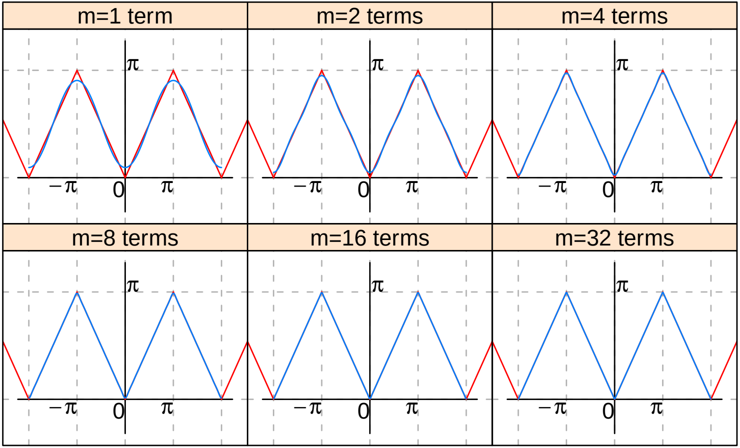
Hence, the Fourier series of $f(x) = |x|$ is

$$\frac{\pi}{2} - \sum_{k=1}^{\infty}\frac{4}{\pi(2k-1)^2}\cos\left((2k-1)x\right) .$$

(example 2, con't)

Let us examine plots of the partial sums to m terms

$$\frac{\pi}{2} - \sum_{k=1}^{m} \frac{4}{\pi(2k-1)^2} \cos\left((2k-1)x\right) .$$



Complex Fourier series I

We have used real-valued functions $\sin(nx)$ and $\cos(nx)$ as our orthonormal system for the linear space E but we can also use complex-valued functions. In this case, we should amend our inner product to

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx\,.$$

A suitable orthonormal system in this case is the collection of functions

$$\left\{1, e^{ix}, e^{-ix}, e^{i2x}, e^{-i2x}, \ldots\right\}\,.$$

Then if $f \in E$ we have a representation, known as the complex Fourier series of $f \in E$, given by

$$\sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx, \quad n = 0, \pm 1, \pm 2, \ldots\,.$$

Complex Fourier series II

Euler's formula ($e^{ix} = \cos(x) + i\sin(x)$) gives for $n = 1, 2, \ldots$ that

$$\begin{aligned} e^{inx} &= \cos(nx) + i\sin(nx) \\ e^{-inx} &= \cos(nx) - i\sin(nx) \end{aligned}$$

and $e^{i0x} = 1$. Using these relations it can be shown that for $n = 1, 2, \ldots$

$$c_n = \frac{a_n - ib_n}{2}, \qquad c_{-n} = \frac{a_n + ib_n}{2}\,.$$

Hence,

$$a_n = c_n + c_{-n}, \qquad b_n = i(c_n - c_{-n})$$

and

$$c_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-i0x}dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)dx = \frac{a_0}{2}\,.$$

Fourier transforms

Introduction

- We have seen how functions $f : [-\pi, \pi] \to \mathbb{C}, f \in E$ can be studied in alternative forms using closed orthonormal systems such as

$$\sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx \quad n = 0, \pm 1, \pm 2, \dots .$$

The domain $[-\pi, \pi]$ can be swapped for a general interval $[a, b]$ and the function can be regarded as L-periodic and defined for all $\mathbb{R}$, where $L = (b - a) < \infty$ is the length of the interval.

- We shall now consider the situation where $f : \mathbb{R} \to \mathbb{C}$ may be a non-periodic function.

Fourier transform

Definition (Fourier transform)

For $f : \mathbb{R} \to \mathbb{C}$ define the Fourier transform of f to be the function $F : \mathbb{R} \to \mathbb{C}$ given by

$$F(\omega) = \mathcal{F}_{[f]}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

whenever the integral exists.

Note two key changes from the Fourier series, now that the function $f(x)$ is no longer constrained to be periodic:

1. the bounds of integration are now $[-\infty, \infty]$ instead of $[-\pi, \pi]$, since the function's "period" is now unbounded – it is aperiodic.
2. the frequency parameter inside the complex exponential previously took only integer values n, but now it must take all real values ω.

We shall use the notation $F(\omega)$ or $\mathcal{F}_{[f]}(\omega)$ as convenient, and refer to it as "the representation of $f(x)$ in the frequency (or Fourier) domain."

For functions $f : \mathbb{R} \to \mathbb{C}$ define the two properties

1. piecewise continuous: if f is piecewise continuous on every finite interval. Thus f may have an infinite number of discontinuities but only a finite number in any subinterval.
2. absolutely integrable: if

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty$$

Let $G(\mathbb{R})$ be the collection of all functions $f : \mathbb{R} \to \mathbb{C}$ that are piecewise continuous and absolutely integrable.

Immediate properties

It may be shown that $G(\mathbb{R})$ is a linear space over the scalars $\mathbb{C}$ and that for $f \in G(\mathbb{R})$

1. $F(\omega)$ is defined for all $\omega \in \mathbb{R}$
2. F is a continuous function
3. $\lim_{\omega \to \pm\infty} F(\omega) = 0$

These properties affirm the existence and nice behaviour of the Fourier transform of all piecewise continuous and absolutely integrable functions $f : \mathbb{R} \to \mathbb{C}$. Soon we will see many further properties that relate the behaviour of $F(\omega)$ to that of $f(x)$, and specifically the consequences for $F(\omega)$ when $f(x)$ is manipulated in certain ways.

Example

For $a > 0$, let $f(x) = e^{-a|x|}$ then

$$\begin{aligned}
F(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-a|x|}e^{-i\omega x}dx \\
&= \frac{1}{2\pi}\left\{\int_0^{\infty} e^{-ax}e^{-i\omega x}dx + \int_{-\infty}^{0} e^{ax}e^{-i\omega x}dx\right\} \\
&= \frac{1}{2\pi}\left\{-\left[\frac{e^{-(a+i\omega)x}}{a+i\omega}\right]_0^{\infty} + \left[\frac{e^{(a-i\omega)x}}{a-i\omega}\right]_{-\infty}^{0}\right\} \\
&= \frac{1}{2\pi}\left\{\frac{1}{a+i\omega} + \frac{1}{a-i\omega}\right\} \\
&= \frac{a}{\pi(a^2+\omega^2)}\,.
\end{aligned}$$

Properties

Several properties of the Fourier transform are very helpful in calculations.

First, note that by the linearity of integrals we have that if $f, g \in G(\mathbb{R})$ and $a, b \in \mathbb{C}$ then

$$\mathcal{F}_{[af+bg]}(\omega) = a\mathcal{F}_{[f]}(\omega) + b\mathcal{F}_{[g]}(\omega)$$

and $af + bg \in G(\mathbb{R})$.

Secondly, if f is real-valued then

$$F(-\omega) = \overline{F(\omega)}\,.$$

This property is called Hermitian symmetry: the Fourier transform of a real-valued function has even symmetry in its real part and odd symmetry in its imaginary part. An obvious consequence is that when calculating the Fourier transform of a real-valued function, we need only consider positive values of ω since $F(\omega)$ determines $F(-\omega)$ by conjugacy.

Even and odd real-valued functions

Theorem

If $f \in G(\mathbb{R})$ is an even real-valued function then F is even and purely real-valued. If f is an odd real-valued function then F is odd and purely imaginary.

Proof.

Suppose that f is even and real-valued then

$$\begin{aligned} F(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\left[\cos(\omega x) - i\sin(\omega x)\right]dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)\cos(\omega x)dx\,. \end{aligned}$$

Hence, F is real-valued and even (the imaginary part has vanished and both f and $\cos(\omega x)$ are themselves even functions). The second part follows similarly. □

Shift and scale properties

Theorem

Let $f \in G(\mathbb{R})$ *and* $a, b \in \mathbb{R}$ *with* $a \neq 0$ *and define*

$$g(x) = f(ax + b)$$

then $g \in G(\mathbb{R})$ *and*

$$\mathcal{F}_{[g]}(\omega) = \frac{1}{|a|} e^{i\omega b/a} \mathcal{F}_{[f]}\left(\frac{\omega}{a}\right)$$

Proof

Set $y = ax + b$ so for $a > 0$ then

$$\mathcal{F}_{[g]}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) e^{-i\omega(\frac{y-b}{a})} \frac{dy}{a}$$

and for $a < 0$

$$\mathcal{F}_{[g]}(\omega) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) e^{-i\omega(\frac{y-b}{a})} \frac{dy}{a} .$$

Hence,

$$\mathcal{F}_{[g]}(\omega) = \frac{1}{|a|} e^{i\omega b/a} \frac{1}{2\pi} \int_{-\infty}^{\infty} f(y) e^{-i\omega y/a} dy = \frac{1}{|a|} e^{i\omega b/a} \mathcal{F}_{[f]}\left(\frac{\omega}{a}\right) .$$

□

Special cases

Two special cases are worth highlighting.

1. Suppose that $b = 0$ so $g(x) = f(ax)$ and so

$$\mathcal{F}_{[g]}(\omega) = \frac{1}{|a|}\mathcal{F}_{[f]}\left(\frac{\omega}{a}\right) .$$

2. Suppose that $a = 1$ so $g(x) = f(x + b)$ and so

$$\mathcal{F}_{[g]}(\omega) = e^{i\omega b}\mathcal{F}_{[f]}(\omega) .$$

Theorem

For $f \in G(\mathbb{R})$ and $c \in \mathbb{R}$ then

$$\mathcal{F}_{[e^{icx} f(x)]}(\omega) = \mathcal{F}_{[f]}(\omega - c)\,.$$

Proof.

$$\begin{aligned}
\mathcal{F}_{[e^{icx} f(x)]}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{icx} f(x) e^{-i\omega x}\,dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x) e^{-i(\omega - c)x}\,dx \\
&= \mathcal{F}_{[f]}(\omega - c)\,.
\end{aligned}$$

□

Note the symmetry (sometimes called a "duality") between the last two properties: a shift in $f(x)$ by b causes $\mathcal{F}_{[f]}(\omega)$ to be multiplied by $e^{i\omega b}$; whereas multiplying $f(x)$ by e^{icx} causes $\mathcal{F}_{[f]}(\omega)$ to be shifted by c.

Modulation property

Theorem

For $f \in G(\mathbb{R})$ and $c \in \mathbb{R}$ then

$$\mathcal{F}_{[f(x)\cos(cx)]}(\omega) = \frac{\mathcal{F}_{[f]}(\omega - c) + \mathcal{F}_{[f]}(\omega + c)}{2}$$
$$\mathcal{F}_{[f(x)\sin(cx)]}(\omega) = \frac{\mathcal{F}_{[f]}(\omega - c) - \mathcal{F}_{[f]}(\omega + c)}{2i}.$$

Proof.

We have that

$$\begin{aligned}
\mathcal{F}_{[f(x)\cos(cx)]}(\omega) &= \mathcal{F}_{\left[f(x)\frac{e^{icx}+e^{-icx}}{2}\right]}(\omega) \\
&= \frac{1}{2}\mathcal{F}_{[f(x)e^{icx}]}(\omega) + \frac{1}{2}\mathcal{F}_{[f(x)e^{-icx}]}(\omega) \\
&= \frac{\mathcal{F}_{[f]}(\omega - c) + \mathcal{F}_{[f]}(\omega + c)}{2}.
\end{aligned}$$

Similarly, for $\mathcal{F}_{[f(x)\sin(cx)]}(\omega)$. □

A major application of the modulation property

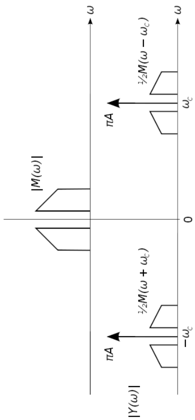
The last two theorems are the basis for broadcast telecommunications that encode and transmit using **a**mplitude **m**odulation of a carrier (e.g. "AM radio"), for receivers that decode the AM signal using a tuner.

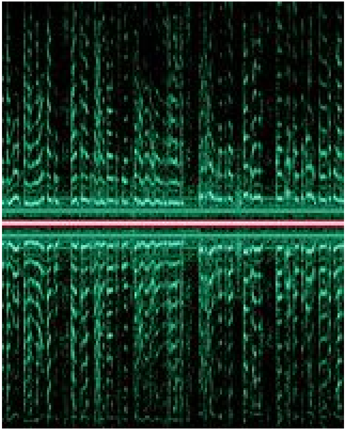
Radio waves propagate well through the atmosphere in a frequency range (or "spectrum") measured in the gigaHertz, with specific bands allocated by government for commercial broadcasting, mobile phone operators, etc. A band around 1 megaHertz (0.3 to 3.0 MHz) is allocated for AM radio, and a band around 1 gigaHertz (0.3 to 3.0 GHz) for mobile phones, etc.

A human audio signal $f(t)$ occupies less than 10 kHz, but its spectrum $F(\omega)$ is shifted up into the MHz or GHz range by multiplying the sound waveform $f(t)$ with a carrier wave e^{ict} of frequency c, yielding $F(\omega - c)$. Its band**width** remains 10 kHz, so many many different channels can be allocated by choices of c. The AM signal received is then multiplied by e^{-ict} in the tuner, shifting its spectrum back down by c, restoring $f(t)$.

This ("single sideband" or SSB) approach requires a complex carrier wave e^{ict}. Devices can be simplified by using a purely real carrier wave $\cos(ct)$, at the cost of shifting in both directions $F(\omega - c)$ and $F(\omega + c)$ as noted, doubling the bandwidth and power requirements.

Example of double-sideband modulation in AM broadcasting





Left: Double-sided spectra of baseband and (modulated) AM signals.
Right: Spectrogram (frequency spectrum versus time) of an AM broadcast shows its two sidebands (green), on either side of the central carrier (red).

Derivatives

There are further properties relating to the Fourier transform of derivatives that we shall state here but omit further proofs.

Theorem

If f is such that both $f, f' \in G(\mathbb{R})$ then

$$\mathcal{F}_{[f']}(\omega) = i\omega \mathcal{F}_{[f]}(\omega)\,.$$

It follows by concatenation that for n^{th}-order derivatives $f^{(n)} \in G(\mathbb{R})$

$$\mathcal{F}_{[f^{(n)}]}(\omega) = (i\omega)^n \mathcal{F}_{[f]}(\omega)\,.$$

In Fourier terms, taking a derivative (of order n) is thus a kind of filtering operation: the Fourier transform of the original function is just multiplied by $(i\omega)^n$, which emphasizes the higher frequencies while discarding the lower frequencies.

The notion of derivative can thus be generalized to non-integer order, $n \in \mathbb{R}$ instead of just $n \in \mathbb{N}$. In fields like fluid mechanics, it is sometimes useful to have the 0.5^{th} or 1.5^{th} derivative of a function, $f^{(0.5)}$ or $f^{(1.5)}$.

Application of the derivative property

In a remarkable way, the derivative property converts calculus problems (such as solving differential equations) into much easier algebra problems. Consider for example a 2^{nd}-order differential equation such as

$$af''(x) + bf'(x) + cf(x) = g(x)$$

where $g(x)$ is some known function or numerically sampled behaviour whose Fourier transform $G(\omega)$ is known or can be computed. Solving this common class of differential equation requires finding the function $f(x)$ for which the equation is satisfied. How can this be done?

By taking Fourier transforms of both sides of the differential equation and applying the derivative property, we immediately get a simple algebraic equation in terms of $G(\omega) = \mathcal{F}_{[g]}(\omega)$ and $F(\omega) = \mathcal{F}_{[f]}(\omega)$:

$$[a(i\omega)^2 + bi\omega + c]F(\omega) = G(\omega)$$

Now we can express the Fourier transform of our desired solution $f(x)$

$$F(\omega) = \frac{G(\omega)}{-a\omega^2 + bi\omega + c}$$

and wish that we could "invert" $F(\omega)$ to express $f(x)$!

Inverse Fourier transform

There is an inverse operation for recovering a function f given its Fourier transform $F(\omega) = \mathcal{F}_{[f]}(\omega)$ which takes the form

$$f(x) = \int_{-\infty}^{\infty} \mathcal{F}_{[f]}(\omega) e^{i\omega x} d\omega\,,$$

which you will recognize as the property of an orthonormal system in the space of continuous functions, using the complex exponentials $e^{i\omega x}$ as its basis elements.

More precisely, the following holds.

Theorem (Inverse Fourier transform)

If $f \in G(\mathbb{R})$ then for every point $x \in \mathbb{R}$ where the one-sided derivatives exist,

$$\frac{f(x-) + f(x+)}{2} = \lim_{M \to \infty} \int_{-M}^{M} \mathcal{F}_{[f]}(\omega) e^{i\omega x} d\omega\,.$$

Convolution

An important operation between two functions in signal processing, and in many other applications, is convolution defined as follows.

Definition (Convolution)

If f and g are two functions $\mathbb{R} \to \mathbb{C}$ then the convolution operation, written $f * g$, creating a third function, is given by

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy$$

whenever the integral exists.

Exercise: show that the convolution operation is commutative, that is $f * g = g * f$.

Fourier transforms and convolutions

The importance of Fourier transform techniques in signal processing rests, in part, on the following result that leads to much simpler descriptions and mathematical formulae in the Fourier domain.

Theorem (Convolution theorem)

For $f, g \in G(\mathbb{R})$ then

$$\mathcal{F}_{[f*g]}(\omega) = 2\pi \mathcal{F}_{[f]}(\omega) \cdot \mathcal{F}_{[g]}(\omega)\,.$$

The convolution integral, whose definition explicitly required integrating the product of two functions for all possible relative shifts between them, to generate a new function in the variable of the amount of shift, is now seen to correspond to the much simpler operation of multiplying together both of their Fourier transforms.

Proof

We have that

$$
\begin{aligned}
\mathcal{F}_{[f*g]}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}(f*g)(x)e^{-i\omega x}dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x-y)g(y)dy\right)e^{-i\omega x}dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x-y)e^{-i\omega(x-y)}g(y)e^{-i\omega y}dxdy \\
&= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} f(x-y)e^{-i\omega(x-y)}dx\right)g(y)e^{-i\omega y}dy \\
&= \mathcal{F}_{[f]}(\omega)\int_{-\infty}^{\infty} g(y)e^{-i\omega y}dy \\
&= 2\pi\mathcal{F}_{[f]}(\omega)\cdot\mathcal{F}_{[g]}(\omega)\,.
\end{aligned}
$$

$\square$

Some signal processing applications

We can now develop some important concepts and relationships, leading to the remarkable Shannon sampling theorem (the exact representation of continuous functions from mere samples of them at periodic points).

We first note two types of limitations on functions.

Definition (Time-limited)

A function f is time-limited if

$$f(x) = 0 \quad \text{for all } |x| \geq M$$

for some constant M.

Definition (Band-limited)

A function $f \in G(\mathbb{R})$ is band-limited if

$$\mathcal{F}_{[f]}(\omega) = 0 \quad \text{for all } |\omega| \geq L$$

for some constant L.

Let us first calculate the Fourier transform of the "unit pulse":

$$f(x) = \begin{cases} 1 & a \leq x \leq b \\ 0 & \text{otherwise}. \end{cases}$$

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx = \frac{1}{2\pi} \int_{a}^{b} e^{-i\omega x} dx .$$

So, for $\omega \neq 0, F(\omega) = \left[\frac{1}{2\pi} \left(\frac{e^{-i\omega x}}{-i\omega} \right) \right]_a^b = \frac{e^{-i\omega a} - e^{-i\omega b}}{2\pi i\omega}$

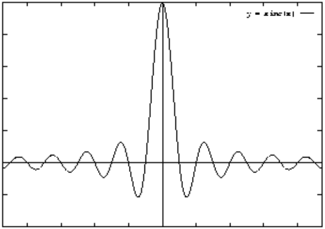
For $\omega = 0$ we have that $F(0) = \frac{1}{2\pi} \int_a^b dx = \frac{(b-a)}{2\pi}$. For the special case when $a = -b$ with $b > 0$ (a zero-centred unit pulse), then

$$F(\omega) = \begin{cases} \frac{e^{i\omega b} - e^{-i\omega b}}{2\pi i\omega} = \frac{\sin(\omega b)}{\omega\pi} & \omega \neq 0 \\ \frac{b}{\pi} & \omega = 0 \end{cases}$$

This important wiggly function, the Fourier transform of the unit pulse, is called a sinc function.

On the previous slide, the sinc was a function of frequency. But a sinc function of x is also important, because if we wanted to strictly low-pass filter a signal, then we would convolve it with a sinc function whose “frequency parameter” corresponds to the cut frequency.

The sinc function plays an important role in the Sampling Theorem, because it allows us to know exactly what a (strictly low-pass) signal does even between the points at which we have sampled it. (This is rather amazing; it sounds like something impossible!)



Note from the functional form that it has periodic zero-crossings, except at its peak where the interval between zeroes is doubled. Note also that the magnitude of oscillations is damped hyperbolically (as $1/x$).

Remarks on Shannon's sampling theorem

- The theorem says that functions which are strictly band-limited by some upper frequency L (that is, $\mathcal{F}_{[f]}(\omega) = 0$ for $|\omega| > L$) are completely determined just by their values at evenly spaced points a distance $\frac{\pi}{L}$ apart. (Proof given in *Information Theory and Coding.*)
- Moreover, we may recover the function exactly given only its values at this sequence of points. It is remarkable that a countable, discrete sequence of values suffices to determine completely what happens between these discrete samples. The "filling in" is achieved by superimposed sinc functions, weighted by the sample values.
- It may be shown that the sinc functions

$$\frac{\sin(Lx - n\pi)}{Lx - n\pi}$$

for $n \in \mathbb{Z}$ form an orthonormal system with inner product

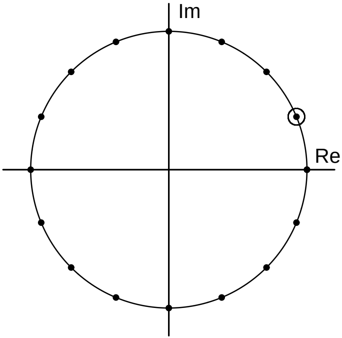
$$\langle f, g \rangle = \frac{L}{\pi} \int_{-\infty}^{\infty} f(x)\overline{g(x)}dx \,.$$

Discrete Fourier Transforms

We now shift attention from functions defined on intervals or on the whole of $\mathbb{R}$ to discrete sequences of values $f[0], f[1], \ldots, f[N-1]$.

An important result in this area of discrete transforms is that the vectors $\{e_0, e_1, \ldots, e_{N-1}\}$ form an orthogonal system in the space $\mathbb{C}^N$ with the usual inner product where the n^{th} component of e_k is given by $(e_k)_n = e^{2\pi ink/N}$ for $n = 0, 1, 2, \ldots, N-1$ and $k = 0, 1, 2, \ldots, N-1$.

The k^{th} vector e_k has N elements and is a discretely sampled complex exponential with frequency k. Its n^{th} element is an N^{th} root of unity, namely the $(nk)^{th}$ power of a primitive N^{th} root of unity:



Applying the usual inner product

$$\langle u, v \rangle = \sum_{n=0}^{N-1} u[n]\overline{v[n]}$$

it may be shown that the squared norm:

$$||e_k||^2 = \langle e_k, e_k \rangle = N\,.$$

In fact, using $\{e_0, e_1, \ldots, e_{N-1}\}$ we can represent any sequence $f = (f[0], f[1], \ldots, f[N-1]) \in \mathbb{C}^N$ by

$$f = \frac{1}{N} \sum_{k=0}^{N-1} \langle f, e_k \rangle e_k\,.$$

Recall the generalized Fourier coefficients that we studied earlier.

Definition (Discrete Fourier Transform/DFT)

The sequence $F[k]$, $k \in \mathbb{Z}$, defined by

$$F[k] = \langle f, e_k \rangle = \sum_{n=0}^{N-1} f[n] e^{-2\pi i n k/N}$$

is called the N-point Discrete Fourier Transform of $f[n]$.

Similarly, for $n = 0, 1, 2, \ldots, N-1$, we have the inverse transform

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} F[k] e^{2\pi i n k/N} .$$

Note that in both of these discrete series defining the Discrete Fourier Transform and its inverse, all of the complex exponential values needed are (nk) powers of a primitive N^{th} root of unity, $e^{2\pi i/N}$. This is the crucial observation that underlies Fast Fourier Transform (FFT) algorithms.

Periodicity

Note that the sequence $F[k]$ has period N since

$$F[k+N] = \sum_{n=0}^{N-1} f[n] e^{-2\pi i n(k+N)/N} = \sum_{n=0}^{N-1} f[n] e^{-2\pi i nk/N} = F[k]$$

using the relation

$$e^{-2\pi i n(k+N)/N} = e^{-2\pi i nk/N} e^{-2\pi i n} = e^{-2\pi i nk/N}.$$

Importantly, note that a complete DFT requires as many (N) Fourier coefficients $F[k]$ to be computed as the number (N) of values in the sequence $f[n]$ whose DFT we are computing.

Properties of the DFT

The DFT satisfies a range of properties similar to those of the FT relating to linearity, and shifts in either the n or k domain.

However, the convolution operation is defined a little differently.

Definition (Cyclical convolution)

The cyclical convolution of two periodic sequences $f[n]$ and $g[n]$ of period N is defined as

$$(f * g)[n] = \sum_{m=0}^{N-1} f[m]g[n-m]\,.$$

Implicitly, because of periodicity, if $[n-m]$ is negative it is taken mod N when only N values are explicit.

It can then be shown that the DFT of $f * g$ is the product $F[k]G[k]$ where F and G are the DFTs of f and g, respectively.

Fast Fourier Transform algorithm

Fast Fourier Transform

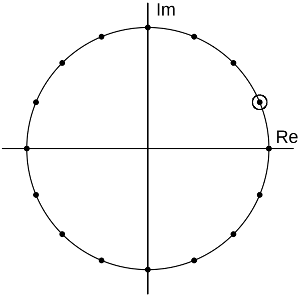
The Fast Fourier Transform is not a new transform but a particular numerical algorithm for computing the DFT.

Since the explicit definition of each Fourier coefficient in the DFT is

$$\begin{aligned} F[k] &= \sum_{n=0}^{N-1} f[n] e^{-2\pi i n k / N} \\ &= f[0] + f[1] e^{-2\pi i k / N} + \cdots + f[N-1] e^{-2\pi i k (N-1)/N} \end{aligned}$$

we can see that in order to compute one Fourier coefficient $F[k]$, using the complex exponential of frequency k, we need to do about $2N$ (complex) additions and multiplications. To compute all N such Fourier coefficients $F[k]$ in this way for $k = 0, 1, 2, \ldots, N-1$ would require about $2N^2$ such operations. Since the number N of samples in a typical audio signal or pixels in an image whose DFT we may need to compute may be $\mathcal{O}(10^6)$, clearly it would be very cumbersome to have to perform $\mathcal{O}(N^2) = \mathcal{O}(10^{12})$ multiplications. Fortunately, very efficient Fast Fourier Transform (FFT) algorithms exist that require instead only $\mathcal{O}(N \log N)$ such operations.

We shall not derive any of the details here but instead give an impression of how such methods operate. Recall that all the multiplications required in the DFT involve the N^{th} roots of unity, and that these in turn are all powers of a primitive N^{th} root of unity $e^{2\pi i/N}$.

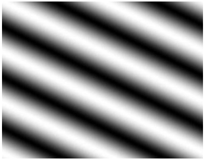


The same points $e^{2\pi ikn/N}$ on the unit circle in the complex plane are used again and again, when the n^{th} value in our sequence $f[n]$ is multiplied by the n^{th} value of a complex exponential $e^{2\pi ikn/N}$ having frequency k, and added together for all n, when computing a DFT coefficient $F[k]$. It is therefore possible to group together common terms associatively and perform far fewer complex multiplications, on such sums of terms.

Extensions to higher dimensions

All of the Fourier methods we have discussed so far have involved only functions or sequences of a single variable. Their Fourier representations have correspondingly also been functions or sequences of a single variable.

But all Fourier techniques can be generalized and apply also to functions of any number of dimensions. For example, images (when pixelized) are discrete two-dimensional sequences $f[n,m]$ giving a pixel value at row n and column m. Their Fourier components are 2D complex exponentials having the form $f[n,m] = e^{2\pi i(kn/N + jm/M)}$ for an image of dimensions NxM pixels, and they have the following "plane wave" appearance with both a "spatial frequency" $\sqrt{k^2 + j^2}$ and an orientation $\arctan(j/k)$:



Similarly, crystallography uses 3D Fourier methods to infer atomic lattice structure from the phases of X-rays scattered by a slowly rotating crystal.

Wavelet Transforms

Wavelets

Wavelets are further bases for representing functions, that have received much interest in both theoretical and applied fields over the past 25 years. They combine aspects of the Fourier (frequency-based) approaches with restored locality, because wavelets are size-specific **local** undulations.

The approach fits into the general scheme of expanding a function $f(x)$ using orthonormal functions. Dyadic transformations of some generating wavelet $\Psi(x)$ spawn an orthonormal wavelet basis $\Psi_{jk}(x)$, for expansions of functions $f(x)$ by doubly-infinite series with wavelet coefficients c_{jk}:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{jk} \Psi_{jk}(x)$$

The wavelets $\Psi_{jk}(x)$ are generated by shifting and scaling operations applied to a single original function $\Psi(x)$, known as the mother wavelet.

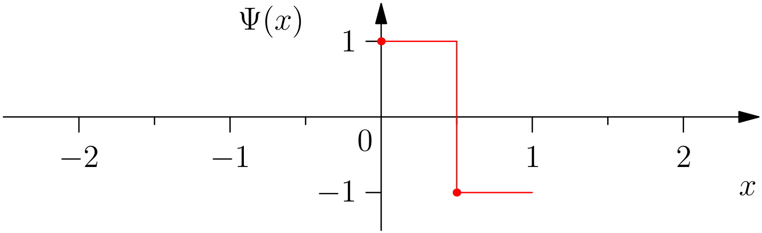
The orthonormal "daughter wavelets" are all dilates and translates of their mother (hence "dyadic"), and are given for integers j and k by

$$\Psi_{jk}(x) = 2^{j/2}\Psi(2^j x - k)$$

The Haar wavelet

An elementary example is the Haar wavelet, whose mother function is both localized and bipolar with a particular scale, defined by

$$\Psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2}\,, \\ -1 & \text{if } \frac{1}{2} \leq x < 1\,, \\ 0 & \text{otherwise}\,. \end{cases}$$



Wavelet dilations and translations

The Haar mother wavelet is localized and has a width (or scale) of 1. The dyadic dilates of $\Psi(x)$, namely,

$$\ldots, \Psi(2^{-2}x), \Psi(2^{-1}x), \Psi(x), \Psi(2x), \Psi(2^2x), \ldots$$

have widths $\ldots, 2^2, 2^1, 1, 2^{-1}, 2^{-2}, \ldots$ respectively.
Since the dilate $\Psi(2^j x)$ has width 2^{-j}, its translates

$$\Psi(2^j x - k) = \Psi(2^j(x - k2^{-j})), \quad k = 0, \pm 1, \pm 2, \ldots$$

cover the whole x-axis. The computed coefficients c_{jk} constitute a Wavelet Transform of the function $f(x)$. There are many different possible choices for the mother wavelet function (besides the Haar), tailored for different purposes. Of course, the wavelet coefficients c_{jk} that result will be different for those different choices of wavelets.

Just as with Fourier transforms, there are fast wavelet implementations that exploit structure. Typically they work in a coarse-to-fine pyramid, with each successively finer scale of wavelets applied to the difference between a down-sampled version of the original function and its full representation by all preceding coarser scales of wavelets.

Interpretation of c_{jk}

How should we intrepret the wavelet coefficients c_{jk}?

Since the Haar wavelet function $\Psi(2^j x - k)$ vanishes except when

$$0 \le 2^j x - k < 1\,, \quad \text{that is} \quad k2^{-j} \le x < (k+1)2^{-j},$$

we see that c_{jk} gives us information about the behaviour of f near the point $x = k2^{-j}$ measured on the scale of 2^{-j}.

For example, the coefficients $c_{(-10,k)}$, $k = 0, \pm 1, \pm 2, \dots$ correspond to variations of f that take place over intervals of length $2^{10} = 1024$, while the coefficients $c_{(10,k)}$ $k = 0, \pm 1, \pm 2, \dots$ correspond to fluctuations of f over intervals of length 2^{-10}.

These observations help explain how wavelet representations extract local structure over many different scales of analysis and can be exceptionally efficient schemes for representing functions. This makes them powerful tools for analyzing signals, compressing images, extracting structure and recognizing patterns.

Properties of naturally arising data

Much naturally arising data is better represented and processed using wavelets, because wavelets are localized and better able to cope with discontinuities and with structures of limited extent. Whereas every Fourier coefficient is computed over the entire extent of the input signal or function (i.e. the bounds of the Fourier integral span the entire input domain), each wavelet has its own local domain, and independent wavelet coefficients are computed for different localities.

Another common aspect of naturally arising data is *self-similarity across scales*, similar to the fractal property. For example, nature abounds with concatenated branching structures at successive size scales. The dyadic generation of wavelet bases mimics this self-similarity.

Finally, wavelets are tremendously good at data compression. This is because they decorrelate data locally: the information is statistically concentrated in just a few wavelet coefficients. The old standard image compression tool JPEG was based on squarely truncated sinusoids. The new JPEG-2000, based on Daubechies wavelets, is a superior compressor.

Case study in image compression: comparison between patchwise Fourier (DCT) and wavelet (DWT) encodings

In 1994, the JPEG Standard was published for image compression using local 2D Fourier transforms (actually discrete cosine transforms [DCT] since images are real, not complex) on small $[8 \times 8]$ tiles of pixels. Each transform produces 64 coefficients and so is not itself a reduction in data.

But because high spatial frequency coefficients can be quantized much more coarsely than low ones for satisfied human perceptual consumption, a quantization table allocates bits to the Fourier coefficients accordingly. The higher frequency coefficients are resolved with fewer bits (often 0).

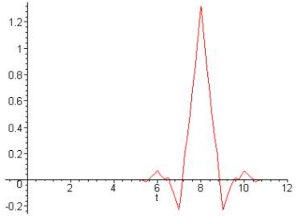
By reading out these quantized Fourier coefficients in a low-frequency to high-frequency sequence, long runs of 0's arise which allow run-length codes (Huffman coding) to be very efficient. $\sim$ 10:1 image compression causes little perceived loss. Both encoding and decoding (compression and decompression) are easily implemented at video frame-rates.

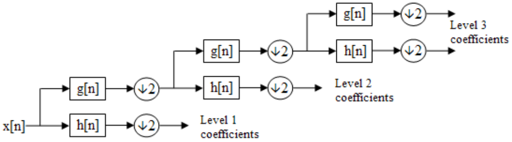
ISO/IEC 10918: *JPEG Still Image Compression Standard.*
JPEG = Joint Photographic Experts Group `http://www.jpeg.org/`

(Image compression case study, continued: DCT and DWT)

Although JPEG performs well on natural images at compression factors below about 20:1, it suffers from visible block quantization artifacts at more severe levels. The DCT basis functions are just square-truncated sinusoids, and if an entire (8×8) pixel patch must be represented by just one (or few) of them, then the blocking artifacts become very noticeable.

In 2000 a more sophisticated compressor was developed using encoders like the Daubechies 9/7 wavelet shown below. Across multiple scales and over a lattice of positions, wavelet inner products with the image yield coefficients that constitute the Discrete Wavelet Transform (DWT): this is the basis of JPEG-2000. It can be implemented by recursively filtering and downsampling the image vertically and horizontally in a scale pyramid.





ISO/IE
15444: *JPEG2000 Image Coding System.* `http://www.jpeg.org/JPEG2000.htm`

Comparing image compressor bit-rates: DCT vs DWT

Whilst a monochrome `.bmp` image assigns 1 byte per pixel and thus has nominally a greyscale resolution of 8 bits per pixel [**8 bpp**], compressed formats deliver much lower **bpp** rates. These are calculated by dividing the total compressed image filesize (in bit count, not bytes) by the total number of pixels in the image. This benchmark image is uncompressed.

Comparing image compressor bit-rates: DCT vs DWT

Left: JPEG compression by 20:1 (Q-factor 10), **0.4 bpp**. The foreground water already shows some blocking artifacts, and some patches of the water texture are obviously represented by a single vertical cosine in an (8×8) pixel block.
Right: JPEG-2000 compression by 20:1 (same reduction factor), **0.4 bpp**. The image is smoother and does not show the blocking quantization artifacts.

Comparing image compressor bit-rates: DCT vs DWT

Left: JPEG compression by 50:1 (Q-factor 3), **0.16 bpp**. The image shows severe quantization artifacts (local DC terms only) and is rather unacceptable.
Right: JPEG-2000 compression by 50:1 (same reduction factor), **0.16 bpp**. At such low bit rates, the Discrete Wavelet Transform gives much better results.

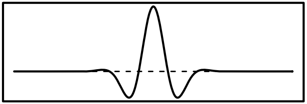
Other classes of wavelets

- Classically, when Yves Meyer gave the original formulation of wavelets ("ondelettes") in a 1985 Bourbaki seminar in Paris, there were 5 strong requirements: the wavelets had all to be dilates and translates of each other, they had to have strictly compact support (equal to 0 outside of some interval), all their derivatives had to exist everywhere, and they had to form an orthonormal basis.
- Today, it is much easier to be wavelet. One of Meyer's students, Stefan Mallat, has said any zero-mean function can be a wavelet.
- In multiple dimensions, we add other transformations based on group theory. For example, for image analysis and vision, we use 2D wavelets that are also rotates of each other in the plane.
- One of the most useful features of wavelets is the ease with which the wavelet functions can be adapted for given scientific problems.
- Many applied fields have started to make use of wavelets, including astronomy, acoustics, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, EEG, ECG, earthquake prediction, radar, etc.

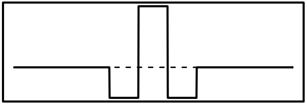
Gabor real and imaginary parts resemble Newton kernels in the calculus

Gabor Wavelets as 1st- and 2nd-order Differential Operators

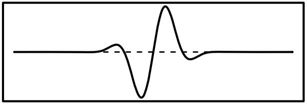
$$\mathbf{Re}\{e^{-x^2}e^{i3x}\} = e^{-x^2}\cos(3x)$$



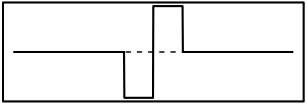
2nd finite difference kernel: $-f''(x_i)$
$\approx -f(x_{i-1}) + 2f(x_i) - f(x_{i+1})$



$$\mathbf{Im}\{e^{-x^2}e^{i3x}\} = e^{-x^2}\sin(3x)$$



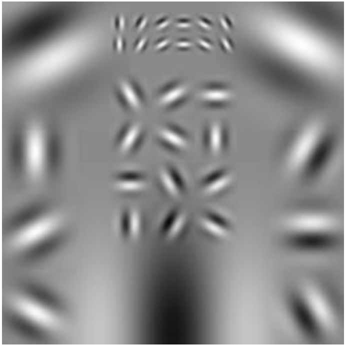
1st finite difference kernel: $f'(x_i)$
$\approx -f(x_i) + f(x_{i+1})$



Wavelets in computer vision and pattern recognition

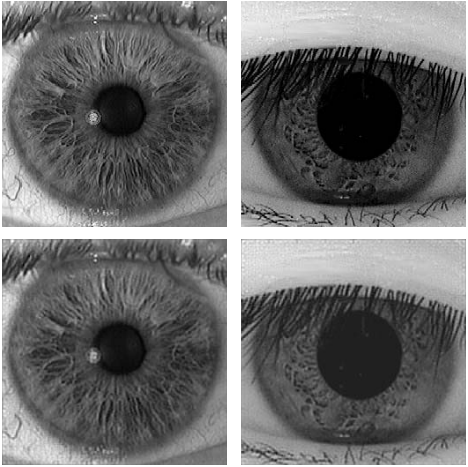
2D Gabor wavelets (defined as a complex exponential plane-wave times a Gaussian windowing function) are extensively used in computer vision.

As multi-scale image encoders, and as pattern detectors, they form a complete basis which can extract image structure with a vocabulary of: location, scale, spatial frequency, orientation, and phase (or symmetry). This collage shows a 4-octave ensemble of such wavelets, differing in size (or spatial frequency) by factors of two, having five sizes, six orientations, and two quadrature phases (even/odd), over a lattice of spatial positions.



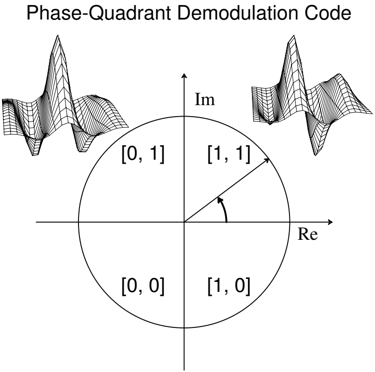
Complex natural patterns are very well represented in such terms.

The upper panels show two iris images (acquired in near-infrared light); caucasian iris on the left, and oriental iris on the right.



The lower panels show the images reconstructed just from combinations of the 2D Gabor wavelets spanning 4 octaves seen in the previous slide.

Gabor wavelets are the basis for Iris Recognition systems

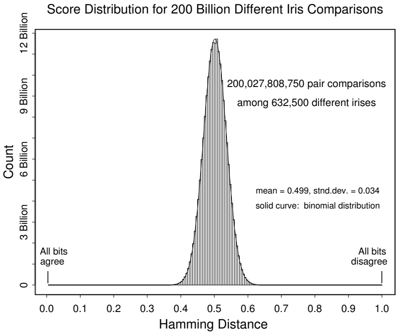


$$h_{Re} = 1 \text{ if } \mathrm{Re} \int_\rho \int_\phi e^{-i\omega(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi \geq 0$$

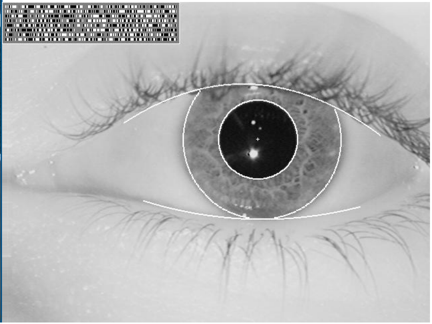
$$h_{Re} = 0 \text{ if } \mathrm{Re} \int_\rho \int_\phi e^{-i\omega(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi < 0$$

$$h_{Im} = 1 \text{ if } \mathrm{Im} \int_\rho \int_\phi e^{-i\omega(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi \geq 0$$

$$h_{Im} = 0 \text{ if } \mathrm{Im} \int_\rho \int_\phi e^{-i\omega(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi < 0$$



Wavelets are much more ubiquitous than you may realize!



At many airports worldwide, the **IRIS** system (Iris Recognition Immigration System) allows registered travellers to cross borders without having to present their passports, or make any other claim of identity. They just look at an iris camera, and (if they are already enrolled), the border barrier opens within seconds. Similar systems are in place for many other applications. The Government of India is currently enrolling the iris patterns of all its 1.2 Billion citizens as a means to access entitlements and benefits (the UIDAI slogan is "To give the poor an identity"), and to enhance social inclusion.