بسم الله الرحمن الرحيم

Sudan University of Science and Technology

College of Graduate Studies

On Fourier Transform and Some Applications

حول تحويل فورير وبعض التطبيقات

A thesis Submitted in partial Fulfillment of the Requirements for

the M.Sc. Degree in Mathematics

submitted by

Azeza Omer Mohamed Mahmoud

Supervised by

Dr.EmadEldeen Abdallah Abdelrahim

August 2016

Dedication

❖To my glorious **Mother**

❖To my dear **Father**

❖ To my dear **Husband**

❖ To my dear **Sisters**

❖ to my dear **brothers**

❖ To my dear **Friends**

Acknowledgement

special thanks to my supervisor, **Dr.EmadEldeen Abdallah Abdelrahim,**

for this outstanding guidance and expert advising ,I would also want to thank any person who support me till search this stage, thanks to all my friends for there support during the year.

Abstract

In this research we present a few model cases of how Fourier series can be applied. The range of applications is large , so our principle of selection has been to choose examples that are both interesting in themselves and have connections with different areas. As an applications we discuss two kinds of examples, the first one isto heat flow, and the second one is application of higher dimensional Fourier series to random walks on a lattice.

الخلاصة

في هذا البحث نقدم نماذج قليله لحالات تطبيق سلسلة فورير ومجال التطبيق واسع جدا لذا فان مبدأ الأختبار قد تم لانتقاء أمثله تكون مشوقه في حد ذاتها ومن ثم ذات صله بالمجالات المختلفه .

في إطار التطبيقات قد ناقشنا نوعين من الأمثله ،النوع الاول هو التطبيق علي التدفق الحراري ، والنوع الثاني هو تطبيق السلوك العشوائي لسلسله فورير ذات الأبعاد الكبيره علي السريان الشبكي .

Table of Contents

Introduction

The Fourier transform converts a set of time domain data vectors into aset of frequency (or per time) domain vectors.

This transform can be effectively used in many applications for example designing and using antennas, image processing and filters, smoothing and sharpening, data processing and analysis and signal and noise estimation .

Our research will be organized as follows:

 In chapter one we study Best L^2 Approximation by Finite Fourier Series, Fourier Series in Action, The first shot in the second industrial revolution, The last shot in the second World War, a nonclassical example: we study the buzz, Convergence of Fourier Series, Rates of convergence and smoothness, Pointwise Convergence vs. Uniform Convergence, Studying Partial Sums via the Dirichlet Kernel: TheBuzz Is Back, The Complex Exponentials Are a Basis for $L^2([0,1])$andMore on the Gibbs Phenomenon

Chapter two illustrative $\boldsymbol{A}*$is Born, we discuss Convolution, really, Properties of Convolution: It's a Lot like Multiplication, Convolution in Action I: A Little Bit on Filtering, Convolution in Action II: differential equations, Probability distributions and probability density functions, The Central Limit Theorem: The Bell Curve Tolls for Theeand Fourier transform formulas under different normalizations.

Chapter three discuss The day of reckoning, The right functions for Fourier Transforms: rapidly decreasing functions, a Very Little on Integrals,distributions, distributions as linear functional, Two important examples of distributions, a Physical analogy for distributions,Limits of distributions and The Fourier Transform of a Tempered distribution.
Chapter four study Fluxions Finis: The End of differential Calculus,Approximations of distributions and Justifying the "Try a FunctionFirst" Principle, The Generalized Fourier Transform Includes the Classical Fourier Transform,Operations on distributions and Fourier Transforms, duality, Changing Signs, Evenness and Oddness,a Function Times a distribution Makes Sense,The derivative Theorem, Shifts and the Shift Theorem, Scaling and the Stretch Theorem,Convolutions and the Convolution Theorem,δ Hard at Work, Filters, redux and diffraction: The sinc function, live and in pure color, with some application.

Chapter (1)

L^2 Approximation, Uniform Convergence and Gibbs Phenomenon

Section(1.1): L^2 Approximation and Fourier Series in Action

We begin this section by studying the BestL^2 Approximation by Finite Fourier Series . Here's a precise statement, and aproof,that a finite Fourier Series of degree N gives the best (trigonometric) approximation of that order in $L^2([0,1])$to a function.

Theorem (1.1.1):

if $f(t)$ is in L^2 ([0,1]) and $\alpha_1, \alpha_2, \dots\dots\dots\alpha_N$are any complex numbers, then

$$\|f - \sum_{n=-N}^{N} (f, e_n)e_n \leq \|f - \sum_{n=-N}^{N} \alpha_n e_n\| .$$

Furthermore,equality holds only when$\alpha_n = (f, e_n)$ for every n.

It's the last statement,on the case of equality ,that leads to the Fourier coefficients in a different

way than solving for them directly as we did originally.

Another way of stating the result is that the orthogonal projection of f onto the

subspace of $L^2([0,1])$ spanned by the $e_n, n = -N, \dots\dots N$ is

$$\sum_{n=-N}^{N} f\,\hat{}(n)e^{\,2\pi int}$$

Here comes the proof.hold on. Write

$$\| f - \sum_{n=-N}^{N} \alpha_n e_n \|^2 = \| f - \sum_{n=-N}^{N} (f, e_n)e_n + \sum_{n=-N}^{N} (f, e_n)e_n - \sum_{n=-N}^{N} \alpha_n e_n \|^2$$

$$= \| \left(f - \sum_{n=-N}^{N} (f, e_n) e_n \right) + \sum_{n=-N}^{N} ((f, e_n) - \alpha_n) e_n \|^2$$

We squared all the norms because we wont to use the properties of inner products to expand the last line.

Using the identity we derived earlier ,the last line equals

$$\| \left(f - \sum_{n=-N}^{N} (f, e_n) e_n \right) + \sum_{n=-N}^{N} ((f, e_n) - \alpha_n) e_n \|^2$$

$$= \| f - \sum_{n=-N}^{N} (f, e_n) e_n \|^2 + 2 Re \left(f - \sum_{n=-N}^{N} (f, e_n) e_n, \sum_{m=-N}^{N} ((f, e_m) - \alpha_m) e_m \right)$$

$$+ \| \sum_{n=-N}^{N} ((f, e_n) - \alpha_n) e_n \|^2$$

This looks complicated ,but the middle term is just asum of multiples of terms of the form

$$\left(f - \sum_{n=-N}^{N} (f, e_n) e_n, e_m \right) = (f, e_m) - \sum_{n=-N}^{N} (f, e_n)(e_n, e_m) = (f, e_m) - (f, e_m)$$
$$= 0,$$

So the whole thing drops out .The final term is

$$\| \sum_{n=-N}^{N} ((f, e_n) - \alpha_n) e_n \|^2 = \sum_{n=-N}^{N} |(f, e_n) - \alpha_n|^2$$

We are left with

$$\| f - \sum_{n=-N}^{N} \alpha_n e_n \|^2 = \| \left(f - \sum_{n=-N}^{N} (f, e_n) e_n \right) \|^2 + \sum_{n=-N}^{N} |(f, e_n) - \alpha_n|^2$$

This completely proves the theorem ,for the right hand side is the sum of two positive terms and hence $\| f - \sum_{n=-N}^{N} \alpha_n e_n \|^2 \geq \| \left(f - \sum_{n=-N}^{N} (f, e_n) e_n \right) \|^2$

With equality holding if and only if

$$\sum_{n=-N}^{N}|(f,e_n)-\alpha_n|^2=0.$$

The latter holds if and only if $\alpha_n = (f,e_n)$ for all n.

The preceding argument may have seemed labor intensive, but it was all algebra based on the properties of the inner product. Imagine trying to write all of it out in terms of integrals.

Now we discuss Fourier series in Action .we've had a barrage of general information

and structure ,and it's time to pass to the particular and put some of these ideas to work .in the following we want to present a few model cases of how Fourier series

can be applied .the range of applications is vast, so my principle of selection has been to choose example that are both interesting in themselves and have connections with different areas.

the first applications are to heat flow ; these are classical , celebrated problems and should be in your storehouse of general knowledge . Another reason for including them is the form that one of the solutions takes as aconvolution integral . we we'll also look briefly at how the differential equation governing heat flow .

the second application is not classical at all ; in fact, it does not fit into L^2 − theory as we laid it out last time. It has to do, on the one hand, with sound synthesis, and on the other. when we do higher dimensional Fourier analysis , we'll have an application of higher dimensional Fourier series to random walks on lattice . it's cool, and, with alittle probability thrown in the analysis of the problem is not beyond what we know to this point, but enough is enough.

the study of how temperature varies over aregion was the first use by Fourier in the 1820's of the method of expanding afunction into aseries of trigonometric function .the physical phenomenon is described ,at least approximately , by apartial differential equation , and Fourier series can be used to write down solutions .

we'll give abrief standard derivation of the differential equation in one spatial dimension ,so the configuration to think of this one- dimensional rod. the argument involves anumber of common but difficult ,practically undefined terms , first among them the term :''heat'', followed closely by term ''temperature''.

As it is usually stated, heat is atransfer of ''energy''(another undefined term) due to temperature difference; The transfer process is called ''heat''. What gets transferred is energy. Because of this, heat is usually identified as a form of energy and has units of energy. we talk of heat as a' transfer of energy ' ,and hence of 'heat flow' ,because, like so many other physical quantities heat is only interesting if it's associated with achange.

Temperature, more properly called ''thermodynamic temperature'' (for merly "absolute temperature") , is a derived quantity . the temperature of a substance is proportional to the kinetic energy of the atoms in the substance .A substance at temperature 0 (absolute zero) cannot transfer energy – it's not ''hot''. The principle at work , essentially stated by Newton, is:

A temperature difference between tow substance in contact with each other causes atransfer of energy from the substance of higher temperature to the substance of lower temperature, and that's heat ,or heat flow .No temperature difference , no heat.

Back to the rod. The temperature is a function of both the spatial variable x giving the position along the rod and of the time t. we let $u\,(x,t)$ denote the temperature, and the problem is to find it .the description of heat just above ,with alittle amplification ,is enough to propose apartial differential equation that $u(x,t)$ should satisfy . To derive it, we introduce $q(x,t)$,the amount of heat that (flows) per second at x and t (so $q(x,t)$ is the rate at which energy is transfered at $x\ and\ t$) . Newton's law of cooling says that this is proportional to the gradient of the temperature:

$$q(x,t)\;=\;-k\,u_x(x,t)\,,\quad k>0\,.$$

the reason for the minus sign is that if $u_x(x,t)>0$, i.e.,if the temperature is increasing at x , then the rate at which heat flows at x is negative _from hotter to colder , hence back from x. the constant k can be identified with the reciprocal of ''thermal resistance'' of the

substance. For agiven temperature gradient , the higher the resistance the smaller the heat flow per second , and similarly the smaller the resistance the greater the heat flow per second .and similarly the smaller the resistance the greater the heat flow per second.

As the heat flows from hotter to colder , the temperature rises in the colder part of the substance . the rate at which the temperature rises at $x,$ given by $u_t(x,t)$,is proportional to the rate at which heat ''accumulates'' per unit length . Now $q(x,t)$ is

already arate _ the heat flow per second _ so the rate at which heat accumulates per unit length is the rate in minus the rate out per length ,which is (if the heat is flowing from left to right)

$$\frac{q(x,t) - q(x + \Delta x, t)}{\Delta x}$$

Thus in the limit

$$u_t(x,t) = -k' q_x(x,t), k' > 0.$$

The constant k' can be identified with the reciprocal of the "thermal capacity'' perunit length. Thermal resistance and thermal capacity are not the standard terms, but they can be related to standard terms , specific heat. they are used here because of the similarity of heat flow to electrical phenomena.

Now differentiate the first equation with respect to x to get

$$q_x(x,t) = -ku_{xx}(x,t),$$

and substitute this into the second equation to obtain an equation involving $u(x,t)$alone :

$$u_t(x,t) = kk' u_{xx}(x,t).$$

this is the heat equation .

to summarize, in whatever particular context it's applied, the step for a problem based on the heat equation involves:

1. A region in space.
2. An initial distribution of   temperature on that region.

It's natural to think of fixing one of the variables and letting the other change. Then the solution $u(x,t)$ tells you

1. For each fixed time t how the temperature is distributed on the region.
2. At each fixed point x how the temperature is changing over time.

We want to look at two examples of using Fourier series to solve such a problem:

Heat flow on a circle and, more dramatically , the temperature of the earth. These are nice examples because they show different aspects of how the methods can be applied

and, as mentioned above, they exhibit forms of solutions, especially for the circle problem, of a type.

why a circle, why the earth – and why Fourier methods? Because in each case the function $u(x, t)$ will be periodic in one of the variables. In one case we work with periodicity in space and in the other periodicity in time.

Heating a circle suppose a circle is heated up , not necessarily uniformly. This provides an initial distribution of temperature. Heat then flows around the circle and the temperature changes over time .At any fixed time the temperature must be a periodic function of the position on the circle , for if we specify points on the circle by an angle ⊠,is the same at ɵ and ⊠ $+ 2\pi$, since these are the same points.

We can imagine a circle as an interval with the endpoints identified, say the interval

$0 \leq x \leq 1$, and we let $u(x, t)$ be the temperature as a function of position and time .Our analysis will be simplified if we choose units so the heat equation takes the form

$$u_t = \tfrac{1}{2}\, u_{xx}\, ,$$

that is ,so the constant depending on physical attributes of the wire is ½. The function $u(x, t)$ is periodic in the spatial variable x with period 1, i.e. ,

$u(x + 1, t) = u(x, t)$,and we can try expanding it as a Fourier series with coefficients that depend on time :

$$u(x, t) = \sum_{n=-\infty}^{\infty} c_n(t) e^{2\pi i n x} \, where \, c_n(t) = \int_{0}^{1} e^{-2\pi i n x} u(x, t)\, dx.$$

This representation of $c_n(t)$ as an integral together with the heat equation for $u(x, t)$ will allow us to find $c_n(t)$ explicitly. Differentiate $c_n(t)$ with respect to t by differentiating under the integral sign:

$$c_n(t)' = \int_{0}^{1} e^{-2\pi i n x} u_t(x, t) dx\, ;$$

now using $u_t = \dfrac{1}{2} u_{xx}$ we can write this as

$$c_n(t)' = \int_0^1 \frac{1}{2} e^{-2\pi i n x} u_{xx}(x,t)\,dx$$

and integrate by parts (twice) to get the derivatives of u and put them onto $e^{-2\pi i n x}$ .using the facts that $e^{-2\pi i n} = 1 \; and \; u(0,t) = u(1,t)$

(both of which come in when we plug in the limits of integration when integrating by parts)we get

$$c_n(t)' = \int_0^1 \frac{1}{2} u(x,t) \frac{d^2}{dx^2} e^{-2\pi i n x}\,dx$$

$$= \int_0^1 \frac{1}{2} u(x,t)(-4\pi^2 n^2)\, e^{-2\pi i n x}\,dx$$

$$-2\pi^2 n^2 \int_0^1 u(x,t)\, e^{-2\pi i n x}\,dx = -2\pi^2 n^2 c_n(t).$$

We have found that $c_n(t)$ satisfies a simple ordinary differential equation

$$c_n(t)' = -2\pi^2 n^2 c_n(t),$$

whose solution is

$$c_n(t) = c_n(0) e^{-2} e^{-2\pi^2 n^2 t} .$$

this solution   involves the initial value $c_n(0)$ and, in fact, this initial   value should be,and will be ,incorporated into the formulation of the problem in terms of the initial distribution of heat .

At time $t = 0$ we assume that the temperature $u(x,0)$ is specified by some (periodic) function $f(x)$:

$$u(x,0) = f(x), \quad f(x+1) = f(x) \quad for \; all \; x.$$

then using the integral representation for $c_n(t)$,

$$c_n(0) = \int_0^1 u(x,0)\, e^{-2\pi i n x}\,dx$$

$$= \int_0^1 f(x)\, e^{-2\pi inx} dx = f\hat{}(n),$$

The n-th Fourier coefficient of f thus we can write

$$c_n(t) = f\hat{}(n)e^{-2\pi^2 n^2 t}$$

and the general solution of the heat equation is

$$u(x,t) = \sum_{n=-\infty}^{\infty} f\hat{}(n)e^{-2\pi^2 n^2 t}\, e^{2\pi inx}\,.$$

this is a neat way of writing the solution and we could leave it at that, but for reasons we're about to see it's useful to bring back the integral definition of $f\hat{}(n)$ and write the expression differently.

write the formula for $f\hat{}(n)$ as

$$f\hat{}(n) = \int_0^1 f(y)\, e^{-2\pi iny} dy\,.$$

(don't use x as the variable of integration since it's already in use in the formula for $u(x,t)$.) then

$$u(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t}\, e^{2\pi inx} = \int_0^1 f(y)\, e^{-2\pi iny} dy$$

$$= \int_0^1 \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t}\, e^{2\pi in(x-y)} f(y)\, dy\,,$$

Or, with

$$g(x-y,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t}\, e^{2\pi in(x-y)}$$

we have

$$u(x,t) = \int_0^1 g(x-y,t)f(y)\, dy\,.$$

the function

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t}\, e^{2\pi inx}$$

8

is called Green's function ,or the fundamental solution for the heat equation for a circle . Note that g is a periodic function of period 1 in the spatial variable . the expression for the solution $u(x,t)$ is a convolution integral ,a term you have probably heard from earlier classes, but new here. In words, $u(x,t)$ is given by the convolution of the initial temperature $f(x)$ with Green's function $g(x,t)$. this is a very important fact.

In general, whether or not there is extra time dependence as the present case, the integral

$$\int_0^1 g(x-y)f(y)\,dy$$

Is called the convolution of and g . observe that the integral makes sense only if g is periodic .that is for a given x between 0 and 1 and for y varying from 0 to 1 (as the variable of integration) $x-y$ will assume values outside the interval $[0,1]$. if g were not periodic it wouldn't make sense to consider $g(x-y)$, but the periodicity is just what allows us to do that.

If you know the terminology coming from linear systems, the Green's function $g(x,t)$ is the "impulse response" associated with the linear system "heat flow on a circle", meaning

1- Inputs go in: the initial heat distribution $f(x)$.
2- Outputs come out: the temperature $u(x,t)$.
3- Outputs are given by the convolution of g with the input:

$u(x,t) = \int_0^1 g(x-y,t)f(y)\,dy$.

convolutions occur absolutely everywhere in Fourier analysis and we'll be spending a lot of time with them this quarter. In fact, an important result states that convolutions must occur in relating outputs to inputs for linear time invariant systems.

In our example, as a formula for the solution, the convolution may be interpreted as saying that for each time t the temperature $u(x,t)$ at a point x is a kind of smoothed average of the initial temperature distribution $f(x)$. In other setting a convolution integral may have different interpretations.

Heating the earth, storing your wine .the wind blows , the rain falls ,and the temperature at any particular place on earth changes over the course of a year . let's

agree that the temperature at any particular place on earth changes over the course of a year. so that the temperature at any particular place on earth is roughly a periodic function of time ,where the period is 1 year .what about the temperature x-meters under that particular place? how dose the Temperature depend on x and t?

Fix a place on earth and let $u(x,t)$ denote the temperature x meters underground at time t .we assume again that u satisfies the heat equation ,

$u_t = \frac{1}{2} u_{xx}$ this time we try a

Solution of the form

$u(x,t) = \sum_{n=-\infty}^{\infty} c_n(x) e^{2\pi i n t}$,

reflecting the periodicity in time .

Again we have an integral representation of $c_n(x)$ as a Fourier coefficient ,

$c_n(x) = \int_0^1 u(x,t) e^{-2\pi i n t} dt$,

And again we wont to plug into the heat equation and find a differential equation that the coefficients satisfy . the heat equation involves a second (partial) derivative with respect to the spatial variable x ,so we differentiate c_n twice and differentiate u under the integral sign twice with respect to x :

$c_n{}''(x) = \int_0^1 u_{xx}(x,t) e^{-2\pi i n t} dt$.

Using the heat equation and integrating by parts (once) gives

$$c_n{}''(x) = \int_0^1 2u_t(x,t) e^{-2\pi i n t} dt$$

$$= \int_0^1 4\pi i n\, u(x,t) e^{-2\pi i n t} dt = 4\pi i n\, c_n(x).$$

We can solve this second $-$ order differential equation in x easily on noting that

$$(4\pi i n)^{1/2} = \pm(2\pi|n|)^{1/2}(1 \pm i) ,$$

Where we take $1 + i$ when $n > 0$ and $1 - i$ when $n < 0$. we'll leave it to you to decide that the root to take is $-(2\pi|n|)^{1/2}(1 \pm i)$,thus

$$c_n(x) = A_n e^{-(2\pi|n|)^{\frac{1}{2}}(1\pm i)x}.$$

What is the initial value $A_n = c_n(0)$? Again we assume that at $x = 0$ there is aperiodic function of t that models the Temperature (at the fixed spot on earth) over the course of the year. Call this $f(t)$ .then $u(o, t) = f(t)$, and

$$c_n(0) = \int_0^1 u(0, t) e^{-2\pi int} dt = \hat{f}(n).$$

Our solution is then

$$u(x, t) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{-(2\pi|n|)^{\frac{1}{2}}(1\pm i)x} e^{-2\pi int}.$$

that's not a beautiful expression, but it becomes more interesting if we rearrange the exponentials to isolate the periodic parts (the ones that have an i in them)from the nonperiodic part that remains. The latter is $e^{-(2\pi|n|)^{\frac{1}{2}}x}$. the terms then looke like

$$\hat{f}(n) e^{-(2\pi|n|)^{\frac{1}{2}}x} e^{2\pi int \pm (2\pi|n|)^{\frac{1}{2}}ix}.$$

what's interesting here? The dependence on the depth, x. Each term is damped by the exponential

$$e^{-(2\pi|n|)^{\frac{1}{2}}x}$$

And phase shifted by the amount $(2\pi|n|)^{\frac{1}{2}}x$.

Take a simple case. suppose that the Temperature at the surface $x = 0$ is given just by $\sin 2\pi t$ and that the mean annual Temperature is 0 ,i.e.

$$\int_0^1 f(t) \, dt = \hat{f}(0) = 0.$$

All Fourier coefficients other than the first are zero, and the solution reduces to

$$u(x, t) = e^{-(2\pi)^{\frac{1}{2}}x} \sin(2\pi t - (2\pi)^{1/2}x).$$

take the depth x so that $(2\pi)^{1/2}x = \pi$. Then thereTemperature is damped by $e^{-\pi} = 0,04$quite a bit ,and it is half a period (six months) out of phase with the Temperature at the surface .the Temperature$x$ – meters below stays pretty constant because of the damping , and Because of the phase shift it's cool in the summer and warm in the winter.There's a name for a place like that. It's called a cellar.

Now we study the first shot in the second industrial revolution .Many types of diffusion processes are similar enough in principle to the flow of heat that they are modeled by the heat equation , or variant of the heat equation , and Fourier analysis is often used to find solutions . One celebrated example of this was the paper by William Thomson (later Lord Kelvin): "One the theory of the electric telegraph" published in 1855 in the proceedings of the Royal Society.

The high tech industry of the mid to late 19th century was submarine telegraphy. sharp pulses were sent at one end , representing the dots and dashes of Morse code, and in transit, if the cable was very long and if pulses were sent in too rapid a succession , the pulses were observed to smear out and overlap to the degree that at the receiving end it

Was impossible to resolve them .the commercial success of telegraph transmissions between continents depended on undersea cables reliably handling a large volume of traffic. How should cables be designed? The stakes were high and a quantitative analysis was needed.

A qualitative explanation of signal distortion was offered by Michael Faraday, who was shown the phenomenon by Latimer Clark . Clark ,an official of the Electric and International Telegraph company ,had observed the blurring of signals on the Dutch-Anglo line . Faraday surmised that a cable immersed in water became in effect an enormous capacitor, consisting as it dose of two conductors – the wire and the water – separated by insulating material .when a signal was sent, part of the energy went into charging the capacitor , which took time , and when the signal was finished the capacitor discharged and that took time . the delay associated with both charging and discharging distorted the Signal and caused Signals sent too rapidly to overlap.

Thomson took up the problem in two letters to G. Stokes (of stokes' theorem fame), which become the published paper. We won't follow Thomson's analysis at this point , because , with passage of time , it is more easily understood via Fourier transform rather than Fourier series . However, here are some highlights. Think of the whole cable as a (flexible) cylinder with a wire of radius a along the axis and surrounded by

a layer of insulation of radius b (thus of thickness $b - a$). To model the electrical properties of the cables, Thomson introduced the depending on x and b and ϵ, the permittivity of the insulator .His formula was

$$C = \frac{2\pi\epsilon}{\ln(a/b)}.$$

He also introduced the ''resistance per unit length'', denoting it by K. Imagining the cable as a series of infinitesimal pieces, and using Kirchhoff's circuit law and ohm's law on each piece , he argued that the voltage $v(x,t)$ at a distance x from the end of the cable and at a time t must satisfy the partial differential equation

$$v_t = \frac{1}{KC} v_{xx}.$$

Thomson states: "This equation agrees with the well-known equation of the linear motion of heat in a solid conductor, and various forms of solution which Fourier has given are perfectly adapted for answering practical questions regarding the use of the telegraph wire."

After the fact, the basis of the analogy is that charge diffusing through a cable may be described in the same way as heat through a rod, with a gradient in electric potential replacing gradient of temperature. (keep in mind, however, that the electron was not discovered till 1897 .) Here we see K and C playing the role of thermal and thermal capacity in the derivation of the heat equation.

The result of Thomson's analysis that had the greatest practical consequence was his demonstration that "the time at which the maximum electrodynamics   effect of connecting the battery for an instant" occurs for

$$t_{max} = \frac{1}{6} KCx^2.$$

the number t_{max} is what's needed to understand the delay in receiving the signal. It's fact that the distance from the end of the cable, x, comes in squared that's so important. This means, for example, that the delay in a signal sent along a 1000 mile cable will be 100 times as a large as the delay along a 100 mile cable, and not 10 times as large, as was thought. This was Thomson's "Law of squares."

Thomson's work has been called " The first shot in the second industrial revolution.'' This was when electrical engineering became decidedly

mathematical. His conclusions did not go unchallenged, however .Consider this quote of Edward Whitehouse , chief electrician for the Atlantic Telegraph Company, speaking in 1856

I believe nature knows no such application of this law [the law of squares] and I can only regard it as a function of the schools; a forced and violent application of a principle in physics, good and true under other circumstances, but misapplied here.

Thomson's analysis did not prevail and the first transatlantic cable was built without regard to his specifications . Thomson's said they had to design the cable to make KC small. They thought they could just crank up the power. The continents were joined August 5, 1858, after four previous failed attempts. The first successful sent message was August 16. The cable failed three weeks later. Too high a voltage. They fried it.

Rather later, in 1876, Oliver Heaviside greatly extended Thomson's work by including the effects of induction .He derived a more general differential equation for the voltage $v(x, t)$ in the form

$$v_{xx} = K\,C\,v_t + SC\,v_{tt}\,,$$

where S denotes the inductance per unit length and , as before, K and C denote the resistance and capacitance per unit length . the significance of this equation, is that it allows for solutions that represent propagating waves. It is Heaviside's equation that is now usually referred to as the "telegraph equation".

in the following we illustrate the last shot in the second world war . speaking of high stakes diffusion processes, in the early stages of the theoretical analysis of atomic explosives it was necessary to study the diffusion of neutrons produced by fission as they worked their way through a mass of uranium .

An analysis of this problem was carried out by Robert Serber and some students at Berkeley in the summer of 1942, preceding the opening of the facilities at Los Alamos(where the bulk of the work was done and the bomb was built). they found that the so − called "critical mass" needed for an explosive chain reaction was about 60 kg of U^{235} , arranged in a sphere of radius about 9cm (together with a tamper surrounding the Uranium). A less careful model of how the diffusion works gives a critical mass of 200 kg .As the story goes , in the development of the German bomb project (which predated the American

efforts) , Werner Heisenberg worked with a less accurate model and obtained too high a number for the critical mass.

Now we study A non classical example: What's the buzz?. we model a musical tone as a periodic wave .A pure tone is a single sinusoid , while more complicated tones are sums of sinusoids . The frequencies of the higher harmonics are integer multiples of the fundamental harmonic and the harmonics will typically have different energies .As a model of the most "complete " and "uniform " tone we might take a sum of all harmonics , each sounded with the same energy, say 1. If we further assume that the period is 1 (i.e.,that the fundamental harmonic has frequency 1)then we're looking at the signal

$$f(t) = \sum_{n=-\infty}^{\infty} e^{2\pi i n t} .$$

what does this sound like ?Not very pleasant , depending on your tastes . it's a buzz;

all tones are present and the sum of them together is "atonal". we'd like to hear this sometime , so if any of you can program it I'd appreciate it . of course if you program it then : (1) you'll have to use a finite sum ; (2) you'll have to use a discrete version .in other words , you'll have to come up with the "discrete-time buzz",where what we've written down here is sometimes called the " continuous-time buzz".

The expression for $f(t)$ is not a classical Fourier series in any sense. It does not represent a signal with finite energy and the series dose not converge in L^2 or in any other easily defined sense. Nevertheless, the buzz is an important signal for several reasons. What does it look like in the time domain?

In the first problem set you are asked to find a closed form expression for the partial sum

$$D_N(T) = \sum_{n=-N}^{\infty N} e^{2\pi i n t} .$$

Rather than giving it away , let's revert to the real form . Isolating the

$n = 0$ term and combining positive and negative terms we get

$$\sum_{n=-N}^{\infty N} e^{2\pi int} = 1 + \sum_{n=1}^{\infty N} (e^{2\pi int} + e^{-2\pi int}) = 1 + 2 \sum_{n=-N}^{\infty N} \cos 2\pi nt \ .$$

One thing to note is that the value at the origin is $1 + 2N$; by periodicity this is the value at all the integers, and with a little calculus you can check that $1 + 2N$ is the maximum .It's getting bigger and bigger with N. (what's the minimum, by the way?)

Here are some plots for N = 5, 10, and 20:
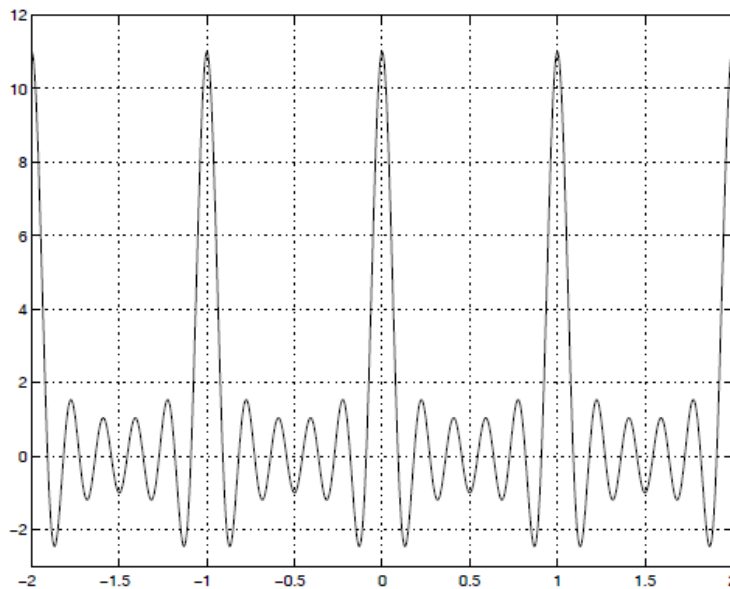
Figure(1.1)
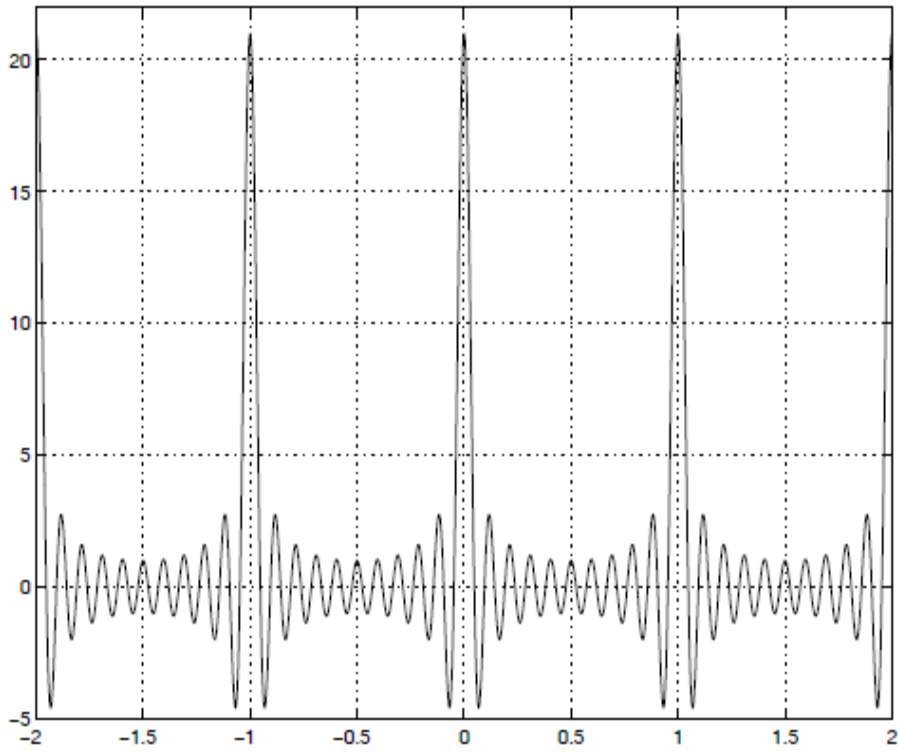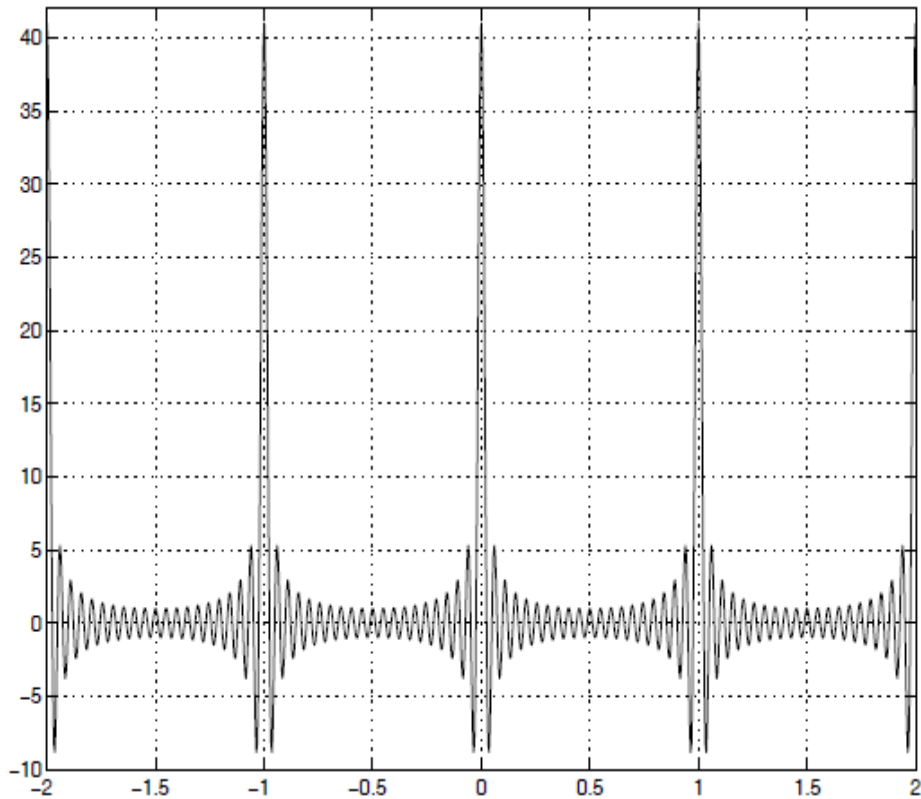
Figure(1.2)

Figure (1.3)

We see that the signal becomes more and more concsentrated at the integers, with higher and higher peaks .in fact, as we'll show later, the sequence of signals $D_N(T)$ tends to a sum of $\delta's$ at the integers as $N \rightarrow \infty$:

17

$$D_N(T) \rightarrow \sum_{n=-\infty}^{\infty} \delta(t-n).$$

In what sense the convergence takes place will also have to wait till later. This all goes to how you that L^2 is not the last word in the development and application of Fourier series.

The sum of regularly spaced δ's is sometimes called an impulse train, and we'll have other descriptive names for it. It is a fundamental object in sampling, the first step in

Turning an analog signal into a digital signal. The finite sum, $D_N(T)$ is called the Dirichlet kernel by mathematicians and it too has a number of applications.

In digital signal processing, particularly music, it's the discrete form of the impulse train --- the discrete time buzz --- that's used. Rather create a sound by adding (sampled) sinusoids one works in the frequency domain and synthesizes the sound from its spectrum. Start with the discrete impulse train, which has all frequencies in equal measure. This is easy to generate. Then shape the spectrum by increasing or decreasing the energies of the various harmonics, perhaps decreasing some to zero. The sound is synthesized from this shaped spectrum, and other operations are also possible.

One final look back at heat. Green's function for the heat equation had the form

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}.$$

At $t \rightarrow 0$ this tends to

$$\sum_{n=-\infty}^{\infty} e^{2\pi i n x},$$

Thecontinuous buzz.

Section (1.2): Uniform Convergence, Complex Exponential and Gibbs Phenomenon

We start by studying the convergence of Fourier Series . the first comment on convergence is – don't go there. Recall that we get tidy mathematical results on convergence of Fourier series if we consider L^2-convergence, or "convergence in mean square". Unpacking the definitions ,that's convergence of the integral of the square of the difference between a function and its finite Fourier series approximation :

$$\lim_{N\to\infty} \int_0^1 |f(t) - \sum_{n=-N}^{N} f\hat{}(n)e^{2\pi i n t} |^2 dt = 0.$$

While this is quite satisfactory in many ways, you might want to know, for computing values of a function , that if you plug a value of t into some finite approximation

$$\sum_{n=-N}^{N} f\hat{}(n)e^{2\pi i n t}$$

You'll be close to the value of the function $f(t)$. and maybe you'd like to know how big you have to take N to get a certain descried accuracy.

All reasonable wishes, but starting to ask about convergence of Fourier series, beyond the L^2- convergence, is starting down a road leading to endless complications, details, and, in the end, probably madness.

Actually- and calmly- for the kinds of functions that come up in applications the answers are helpful and not really so difficult to establish. It's when on inquires into convergence of Fourier series for the most general functions that the trouble really starts. With that firm warning understood, there are a few basic things you ought to know about, if only to know that this can be dangerous stuff.

In the first part of this studying we intention is to summarize the main facts together

With some examples and simple arguments. We'll give careful statements, but we won't enjoy the complete proofs that support them, though in the appendices we'll fill in more of the picture. There we'll sketch the argument for the result at

the heart of the L^2-theory of Fourier series, that the complex exponentials from a basis for $L^2([0, 1])$.

In the following we illustrate how big are the Fourier coefficients?. Suppose that $f(t)$ is square integrable , and let

$$f(t) = \sum_{n=-\infty}^{\infty} f\,\hat{}\,(n)e^{2\pi int}$$

be its Fourier series . Rayleigh's identify says

$$\sum_{n=-\infty}^{\infty} |f\,\hat{}\,(n)|^2 = \int_0^1 |f(t)|^2 dt < \infty.$$

In particular the series

$$\sum_{n=-\infty}^{\infty} |f\,\hat{}\,(n)|$$

Convergence, and it follows that

$$|f\,\hat{}\,(n)|^2 \to o \ as \ n \ \to \ \pm\infty.$$

This is a general result on convergence series − if the series convergence the general term must tend to zero. Knowing that the coefficients tend to zero, can we say how fast?

Here's a simple minded approach that gives some sense of the answer, and shows how answer depends on discontinuities in the function or its derivatives. All of this discussion is based on integration by parts with definite integrals.Suppose, as always, that $f(t)$ is periodic of period 1. By the periodicity condition we have $f(0) = f(1)$ let's assume for this discussion that the function doesn't jump at the endpoints 0 and 1 (like the saw tooth function, below) and that any "problem points" are inside the interval. (this really isn't a restriction. We just want to deal with a single discontinuity for the argument to follow.)that is ,we're imagining that there may be

 trouble at a point $t_0 with \ 0 < t_0 < 1; f(t)$ jumps there, or $f(t)$ is continuous at t_0 but there's a corner, so $f^{\square}(t)$ jumps at t_0, and so on.

Then $n − th$ Fourier coefficient is given by

$$\hat{f}(n) = \int_0^1 e^{-2\pi i n t} f(t) dt .$$

To analyze the situation near t_0 write this as the sum of two integrals:

$$\hat{f}(n) = \int_0^{t_0} e^{-2\pi i n t} f(t) dt + \int_{t_0}^1 e^{-2\pi i n t} f(t) dt .$$

Apply integration by parts to each of these integrals. In doing so, we're going to suppose that at least away from t_0 the function has as many derivatives as we want. Then, on a first pass,

$$\int_0^{t_0} e^{-2\pi i n t} f(t) dt = \left[\frac{e^{-2\pi i n t} f(t)}{-2\pi i n} \right]_0^{t_0} - \int_0^{t_0} \frac{e^{-2\pi i n t} f'(t)}{-2\pi i n} dt$$

$$\int_{t_0}^1 e^{-2\pi i n t} f(t) dt = \left[\frac{e^{-2\pi i n t} f(t)}{-2\pi i n} \right]_{t_0}^1 - \int_{t_0}^1 \frac{e^{-2\pi i n t} f'(t)}{-2\pi i n} dt$$

Add these together. Using $f(0) = f(1)$, this results in

$$\hat{f}(n) = \left[\frac{e^{-2\pi i n t} f(t)}{-2\pi i n} \right]_{t_0^-}^{t_0^+} - \int_0^1 \frac{e^{-2\pi i n t} f'(t)}{-2\pi i n} dt ,$$

Where the notation t_0^- and t_0^+ means to indicate we're looking at the value of $f(t)$ as we take left hand and right hand limits at t_0 . If $f(t)$ is continuous at t_0 then the terms in brackets cancel and we're left with just the integral as an expression for $\hat{f}(n)$.

But if $f(t)$ is not continuous at t_0 - if it jumps, for example – then we don't get cancellation, and we expect that the Fourier coefficient will be of order 1/n in magnitude.

Now suppose that $f(t)$ is continuous at t_0 , and integrate by parts a second time. In the same manner as above, this gives

$$\hat{f}(n) = \left[\frac{e^{-2\pi i n t} f'(t)}{(-2\pi i n)^2} \right]_{t_0^-}^{t_0^+} - \int_0^1 \frac{e^{-2\pi i n t} f'(t)}{(-2\pi i n)^2} dt ,$$

If $f'(t)$ is continuous at t_0 then the bracketed part disappears. If $f'(t)$ is not continuous at t_0 , for example if there is a corner at t_0 ,then the terms do not cancel and we expect the Fourier coefficient to be of size $\frac{1}{n^2}$.

We can continue in this way. The rough rule of thumb may be stated as:

1. if $f(t)$ is not continuous then the Fourier coefficients should have terms like $1/n$.

2. if $f(t)$ is differentiable except for corners ($f(t)$ is continuous but $f'(t)$ is not   then the Fourier coefficients should have some terms like $\frac{1}{n^2}$.

3. If $f''(t)$ exists but is not continuous then the Fourier coefficients should have some terms like $\frac{1}{n^3}$.

A discontinuity in $f''(t)$ is harder to visualize; typically it's a discontinuity in the curvature. For example,imagine a curve consisting of an arc of a circle and a line segment tangent to the circle at their endpoints. Something like
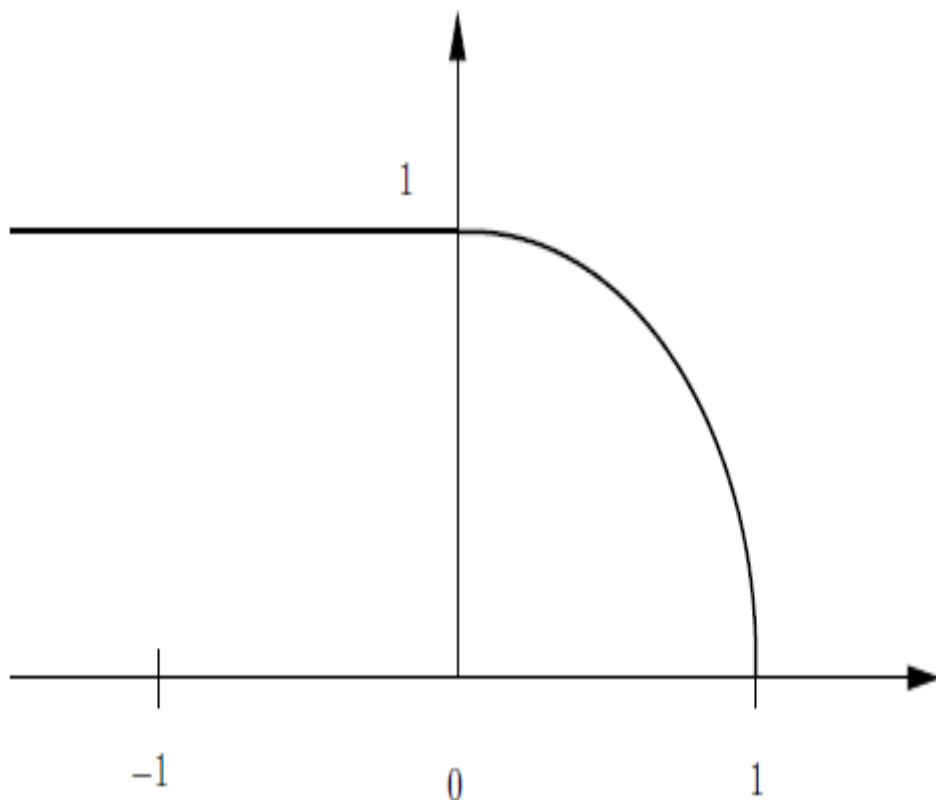

Figure (1.4)

The curve and its first derivative are continuous at the point of tangency, but the second derivative has a jump. If you rode along this path at constant speed you'd feel a jerk – a discontinuity in the acceleration – when you passed through the point of tangency.

Obviously this result extend to discontinuities in higher order derivatives. it also jibes with some examples we had earlier. The square wave

$$f(t) = \left\{ \begin{array}{ll} +1 & 0 \leq t < \frac{1}{2} \\ -1 & \frac{1}{2} \leq t < 1 \end{array} \right.$$

22

Has jump discontinuous, and its Fourier series is

$$\sum_{nodd} \frac{2}{\pi in} e^{2\pi int} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} sin2\pi(2k+1)t.$$

The triangle wave

$$g(t) = \{ \begin{array}{ll} \frac{1}{2}+ t & -\frac{1}{2}\leq t\leq 0 \\ \frac{1}{2}-t & 0 \leq t\leq +\frac{1}{2} \end{array}$$

is continuous but the derivative is discontinuous. (in fact the derivative is the square wave.)Its Fourier series is

$$\frac{1}{4} + \sum_{k=0}^{\infty} \frac{2}{\pi^2(2k+1)^2} \cos(2\pi(2k+t)t).$$

Now we study Rates of convergence and smoothness .The size of the Fourier coefficient tell you something about the rate of convergence of the Fourier series.

There is a precise result on the rate of convergence, which we'll state but not prove:

Theorem (1.2.1):

Suppose that $f(t)$is $p - times$continuously differentiable, where p is at least 1. Then the partial sums

$$S_N(t) = \sum_{n=-N}^{N} f\hat{ }e^{2\pi int}$$

Converge to $f(t)$ point wise and uniformly on $[0,1]$ as $N \to \infty$ .Furthermore

$$max \mid f(t) - S_N(t)\mid \leq constant \frac{1}{N^{P-\frac{1}{2}}}$$

for $0 \leq t \leq 1.$

We won't prove it, but we do want to explain a few things. First, at a meta level, this result has to do with how local properties of the function are reflected in global properties of its Fourier series. In the present setting, "local properties" of a function refers to how smooth it is, i.e., how many times it's continuously

differentiable. About the only kind of "global question" one can ask about series is how fast they converge,at that's what estimated here .The essential point is that the error in the approximation (and indirectly the rate at which the coefficients decrease) is governed by the smoothness (the degree of differentiability) of the signal. The smoother the function – a "local" statement – the better the approximation, and this is not just in the mean, L^2 sense, but uniformly over the interval – a "global" statement.

Let me explain the two terms "pointwise" and "uniformly"; the first is what you think you'd like, but the second is better. "pointwise" convergence means that if you plug in a particular value of t the series converges at the point to the value of the signal at the point. "uniformly" means that the rate at which series converges is the same point in [0, 1]. There are several ways of rephrasing this.

Analytically, the way of capturing the property of uniformly is by making a statement, as above, on the maximum amount the function $f(t)$ can differ from its sequence of approximation $S_N(t)$ for any t in the interval. The "constant" in the inequality will depend on f (typically the maximum of some derivative of some order over the interval, which regulates how much the function wiggles) but not on t – that's uniformity. A geometric picture of uniform convergence may be clearer. A sequence of functions $f_n(t)$ converges uniformly to a function $f(t)$ if the graphs of the $f_n(t)$ get uniformly close to the graph of $f(t)$.we'll leave that second "uniformly" in the sentence to you to specify more carefully (it would force you to restate the analytic condition) but the picture should be clear.

Interestingly, in proving the theorem it's not so hard to show that the partial sums themselves are converging, and how fast. The trickier part is to show that the sums are converging to the value $f(t)$ of the function at every t. At any rate, the takeaway headline from this is:

If the function is smooth, the Fourier series converges in every sense you could want;L^2, pointwise, uniformly.

Now we discuss convergence if it's not continuous?. say

$$f(t) = \begin{cases} t & 0 \leq t < 1 \\ 0 & otherwise \end{cases}$$
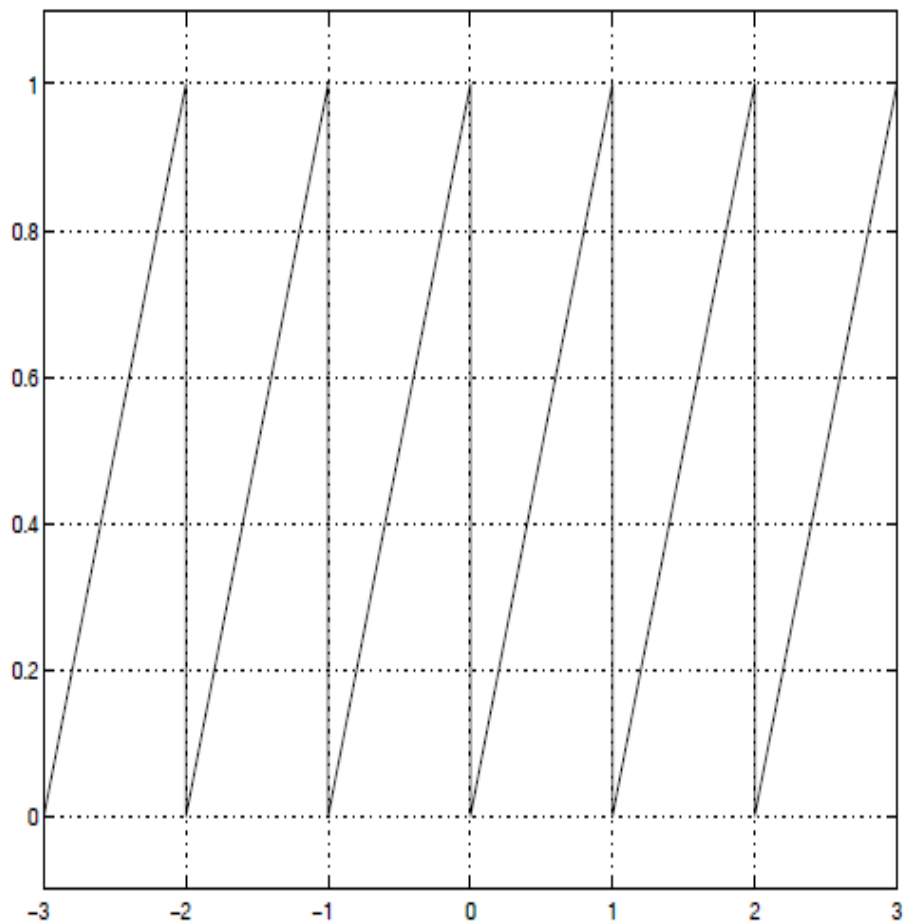
and extended to be periodic of period 1.

Figure (1.5)

The Fourier coefficient are given by

$$f\hat{}(n) = \int_0^1 te^{-2\pi int} dt.$$

Integrating by parts gives, when $n \neq 0,$

$$f\hat{}(n) = [\frac{te^{-2\pi int}}{-2\pi in}]_0^1 - \int_0^1 \frac{1}{-2\pi in} e^{-2\pi int} dt = \frac{i}{2\pi n}$$

$$(use \ 1/i \ = \ -i; \ the \ integral \ is \ 0.)$$

Notice a few things.

1. the coefficients are of the order $1/n$, just as they're supposed to be.
2. the term with $n = 0$ is $1/2$, which we have to get directly, not from the integration by parts step.

(you might also notice the conjugate symmetry in the coefficients, $\hat{f}(-n) = \hat{f}(n)^{\square}$).

So the Fourier series is

$$f(t) = \frac{1}{2} + \sum_{n=-\infty}^{\infty} \frac{i}{2\pi n} e^{2\pi i n t}.$$

Which means that

$$\lim_{N\to\infty} \left\| f(t) - \left(\frac{1}{2} + \sum_{n=-N}^{N} \frac{i}{2\pi n} e^{2\pi i n t} \right) \right\| = 0$$

In the L^2 norm. but what do we get when we plug in a value of t and compute the sum. Even setting aside the obvious problem of adding up an infinite number of terms? Here are the plots for N = 5 , 10 , and 50 of partial sums

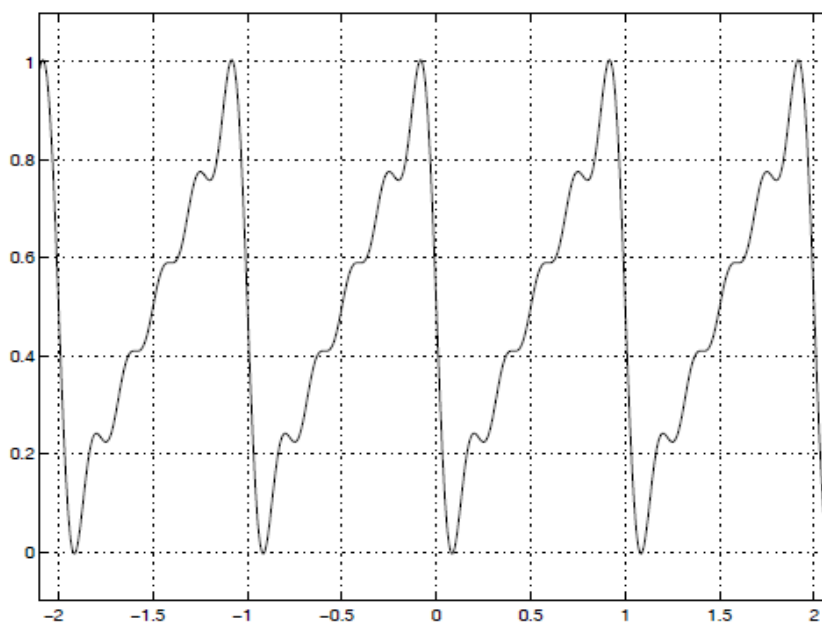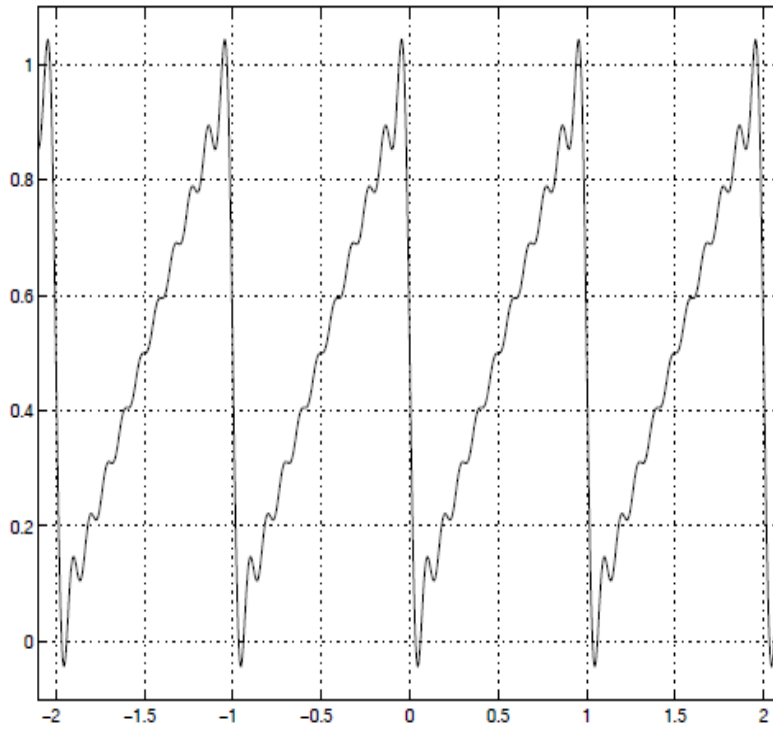$$S_N(t) = \frac{1}{2} + \sum_{n=-N}^{N} \frac{i}{2\pi n} e^{2\pi i n t}.$$
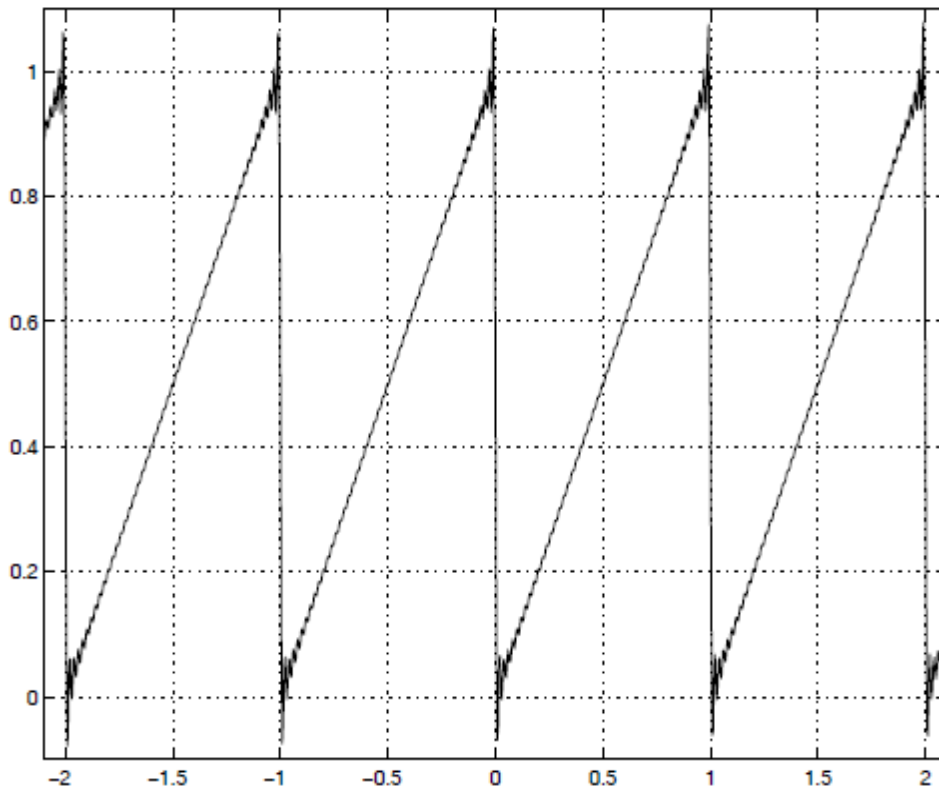

Figure (1.6)

Figure (1.7)

Figure (1.8)

Theorem (1.2.2):

At a jump discontinuity (such as occurs in the sawtooth) the partial sums

$$S_N(t) = \sum_{n=-N}^{N} f\hat{\ } e^{2\pi int}$$

Converge to the converge of the upper and lower values at the discontinuities.

For example, for the sawtooth the partial sums converge to ½ at the points $t = 0, \pm 1, \pm 2, \dots \dots \dots$

Because of this result some people define a value of a function at a jump

Discontinuity to be the average of the upper and lower values. That's reasonable in many contexts – this is one context and we'll see others – but it becomes a religious issue to some so we'll pass without further comment.

We can combine this theorem with the previous theorem to state a useful resultthat's easy to apply in practice:

Theorem (1.2.3): (on pointwise convergence)

suppose that $f(t)$ is continuous with a continuous derivative except at perhaps a finite number of points (in a period). Then for each a $\epsilon[0,1]$,

$S_N(a) \rightarrow$ ½ $(\lim_{t \to a^-} f(t) + \lim_{t \to a^+} f(t))$

as N $\rightarrow \infty$.

If $f(t)$ is continuous at a then the left and right hand limits are equal and we just have$S_n(a) \rightarrow f(a). if f(t)$ has jump at a then we're in the situation in the theorem just above and $S_N(a)$converges to the average of the left and right hand limits.

Thefunny behavior near the corners, where it seems that the approximation are overshooting the signal. Is more interesting. We saw this also with the approximation to the square wave. This is Gibbs phenomenon, named after J.W. Gibbs. it really happens, and it's time to come to terms with it . it was observed experimentally by Michelson and Stratton (that's the same Albert Michelson as in the famous "Michelson and Morley" experiment) who designed a mechanical device to draw finite Fourier series. Michelson and Stratton assumed that the extra wiggles they were seeing at jumps that the was a mechanical problem with

the machine. But Gibbs, who used the sawtooth as an example, showed that the phenomenon is real and does not go away even in the limit. The oscillations many become more compressed, but they don't go away. (However, they do contribute zero in the limit of the L^2 norm of the difference between the function and it's Fourier series.)

a standard way to formulate Gibbs's phenomenon precisely is for a square wave that jumps from $-1\ to\ +1\ at\ t = 0$ when t goes from negative to positive. Away from the single jump discontinuity, $S_N(t)$tends uniformly to the values, +1 or -1 as the case may be, as

N $\rightarrow \infty$. Hence the precise statement of Gibbs's phenomenon will be that the maximum of $S_N(t)$remains greater than 1 as N $\rightarrow \infty$. And that's what is proved:

$$\lim_{N\to\infty} \max S_N(t) = 1.089490 \dots.$$

So the overshoot is almost 9% .

Now, there's something here that may bother you. We have the theorem on pointwise convergence that says at a jump discontinuity the partial sums converge to the average of the values at the jump. We also have Gibbs" phenomenon and the picture of an overshooting oscillation that doesn't go away. How can these two pictures coexist ? if you're confused it's because you're thinking that convergence of $S_N(t)$at, say, $t = 0$ in the sawtooth example, is the same as convergence of the graphs of the $S_N(t)$to the graph of the sawtooth function. But they are not the same thing.

Finally, you should be aware that discontinuities are not at all uncommon. You might introduce jumps via windows or filters, for example. We mentioned earlier that this can be a problem in computer music and images as two-dimensional signals, often have edges. Remember that , a discontinuity or a corner means that you must have infinitely high frequencies in the spectrum, so cutting off the approximation at a certain point is sure to introduce ripples in the computation of values of the function by means of a finite Fourier series approximation.

In the following we illustrate point wise convergence vs. Uniform Convergence. Here's an example, a classic of its type, to show that point wise convergence is not the same as uniform convergence , or what amount to the same thing, that we can have a sequence of function $f_n(t)$with the property that $f_n(t) \rightarrow f(t)$

for every value of t as $n \rightarrow \infty$ but the graphs of the $f_n(t)$ do not ultimately. Look like the graph of $f(t)$. let me describe such a sequence of functions in words, draw a few pictures, and leave it to you .

The $f_n(t)$ will all be defined on $0 \leq t \leq 1$. For each n the graph of the function $f_n(t)$ is zero from $1/n$ to 1 and for $0 \leq t \leq 1/n$ it's an isosceles

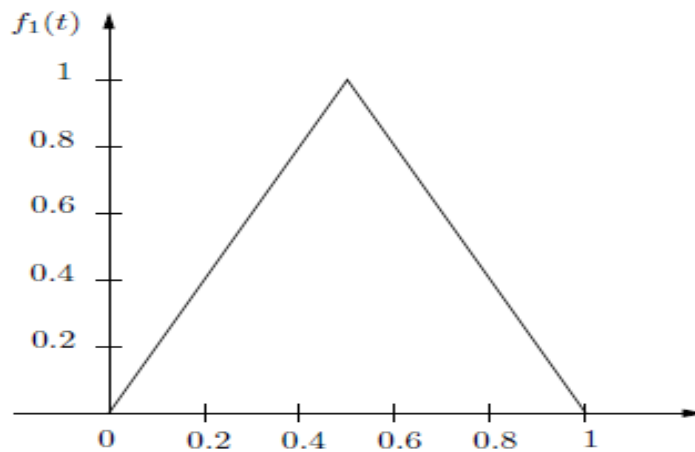triangle with height n^2. Here are pictures of $f_1(t), f_5(t)$ and $f_{10}(t)$.
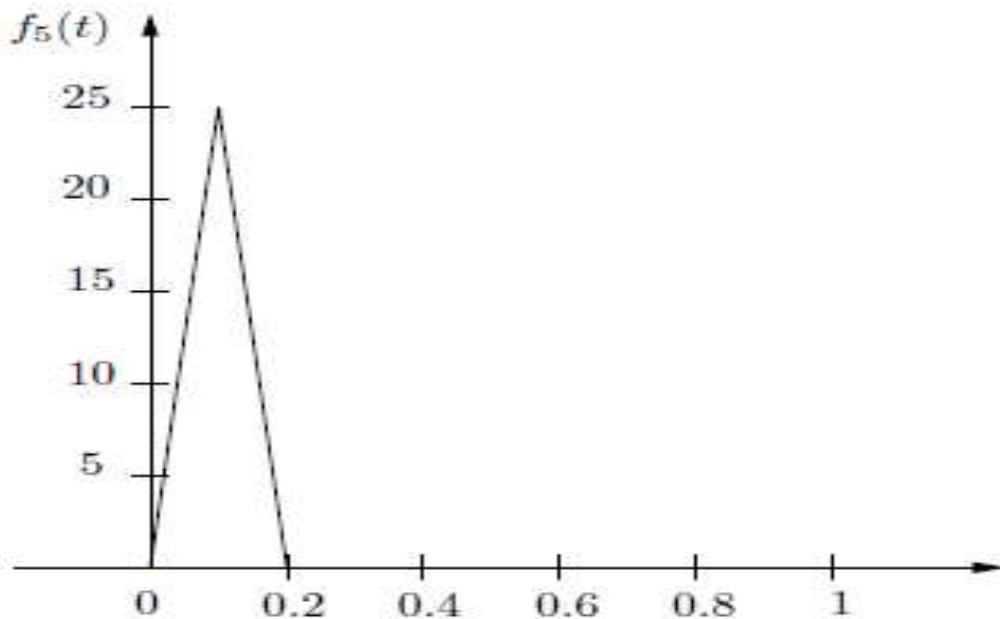

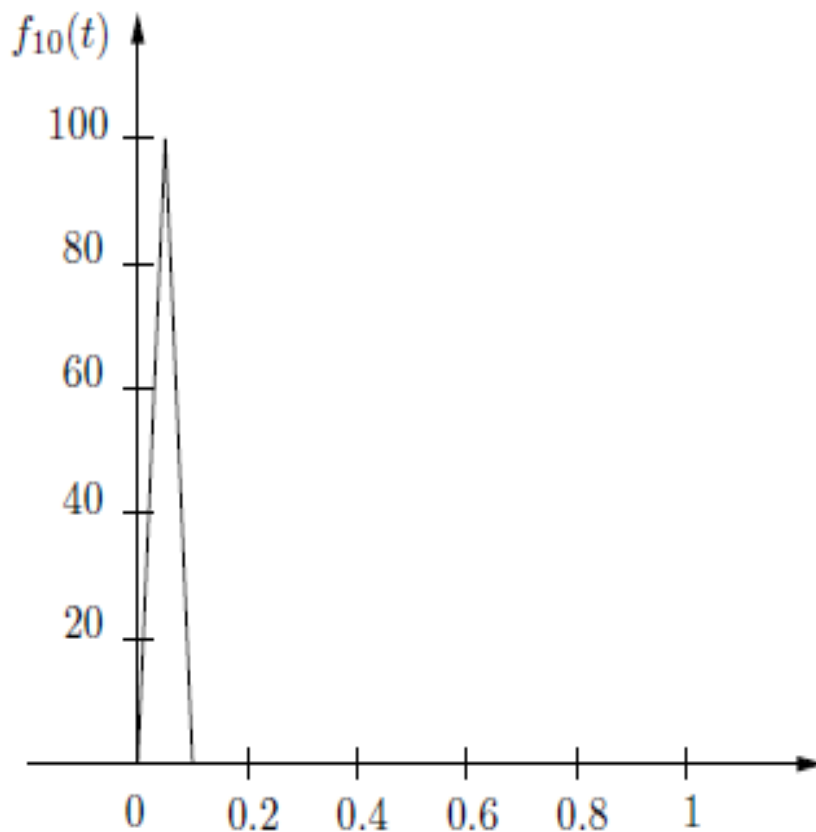**Figure (1.9)**

Figure (1.10)

Figure (1.11)

The peak slides to the left and gets higher and higher as n increases. It's clear that for each t the sequence $f_n(t)$ tends to 0. This is because $f_n(t) = 0$ for all n, and for any $t \neq 0$ eventually, that is, for large enough n, the peak is going to slide to the left of t and $f_n(t)$ will be zero from that n on out. Thus $f_n(t)$ converges point wise to the constant 0. But the graphs of the $f_n(t)$ certainly are not uniformly close to 0.

Now we Studying partial Sums via Dirichlet Kernel: The Buzz Is Back. There are some interesting mathematical tools used to study the partial sums of Fourier series and their convergence properties, as in the theorem we stated earlier on the rate of convergence of the partial sums for p times continuously differentiable functions. In fact, we've already seen the main tool – it's the Dirichlet kernel

$$D_N(t) = \sum_{n=-N}^{N} e^{2\pi i n t}$$

31

We can write a partial sum in what turns out to be helpful way by bringing back in the definition of the Fourier coefficient as an integral.

$$S_N(t) = \sum_{n=-N}^{N} \hat{f}(n)^{2\pi int}$$

$$= \sum_{n=-N}^{N} \left(\int_0^1 f(s)e^{-2\pi ins}\, ds \right) e^{2\pi int}$$

(calling the variable of integration s since we're already using t)

$$= \sum_{n=-N}^{N} \left(\int_0^1 e^{2\pi int} f(s)e^{-2\pi ins}\, ds \right)$$

$$= \int_0^1 \left(\sum_{n=-N}^{N} e^{2\pi int} e^{-2\pi ins} f(s) \right) ds$$

$$= \int_0^1 \left(\sum_{n=-N}^{N} e^{2\pi in(t-s)} \right) f(s)\, ds = \int_0^1 D_N(t-s)f(s)\, ds.$$

Just as we saw in the solution of the heat equation, we have produced a convolution. The integral

$$\int_0^1 D_N(t-s)f(s)\, ds$$

Is the convolution of $f(t)$ with the function $D_N(t)$ and it produces the partial sum $S_N(t)$

Why is this helpful? By means of the convolution integral, estimates for $D_N(t)$ involve both properties of f (on which we make certain assumptions) together with properties of $D_N(t)$, for which we can find an explicit expression. the idea is to view $D_N(t)$ as a geometric series. We can simplify the algebra by factoring out the term corresponding to –N, thereby writing the sum as going from 0 to 2N:

$$\sum_{n=-N}^{N} e^{2\pi i n p} = e^{-2\pi i N p} \sum_{n=0}^{2N} e^{2\pi i n p}$$

$$= e^{-2\pi i N p} \frac{e^{2\pi i (2N+1)p} - 1}{e^{2\pi i p} - 1}$$

(using the formula for the sum of a geometric series $\sum r^n$ with $r = e^{2\pi i p}$)

It's most common to write this in terms of the sine function. Recall that

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

To bring the sine into the expression, above, there's a further little factoring trick that's used often:

$$e^{2\pi i(2N+1)p} - 1 = e^{\pi i (2N+1)p}(e^{\pi i(2N+1)p} - e^{-\pi i(2N+1)p})$$

$$= 2i \, e^{\pi i(2N+1)p} \sin\left(\pi(2N+1)\, p\right)$$

$$e^{2\pi i p} - 1 = e^{\pi i p}(e^{\pi i p} - e^{-\pi i p})$$

$$= 2i \, e^{\pi i p} \sin(\pi p)$$

Therefore

$$e^{-2\pi i N p} \frac{e^{2\pi i(2N+1)p} - 1}{e^{2\pi i p} - 1} = e^{-2\pi i N p} \frac{e^{\pi i(2N+1)p}}{e^{\pi i p}} \frac{2i \sin(\pi(2N+1)\, p)}{2i \sin(\pi p)}$$

$$= \frac{\sin(\pi(2N+1)\, p)}{\sin(\pi p)}.$$

Forming the convolution, as in

$$S_N(t) = \int_0^1 D_N(t-s)f(s)\,ds,$$

Above, shifts the peak at 0 to t, and integrates. The integrand is concentrated around t (as it turns out the peaks at the other integers don't contribute) and in the limit as $N \to \infty$ the integral tends to $f(t)$.

Carrying this out in detail – which we are not going to do – depends on the explicit formula for $D_N(t)$.

The more one assumes about the signal $f(t)$ the more the argument can produce.

This is how one gets the theorem on order of differentiability and rate of convergence of partial sums of Fourier series.

In the following we illustrate the complex exponentials are a basis for $L^2([0,1])$ Remember the second point in our hit parade of the L^2- theory of Fourier series:

The complex exponentials $e^{2\pi int}, n = 0, \pm 1, \pm 2, ...$ from a basis for $L^2([0,1])$, and the partial sums converge to $f(t)$ as N $\rightarrow \infty$ in the L^2-distance. This means that

$$\lim_{N\rightarrow\infty} \left\| \sum_{n=-N}^{N} f\hat{}(n)e^{2\pi int} - f(t) \right\| = 0.$$

We said earlier that we wouldn't attempt a complete proof of this, and we won't. but with the discussion just preceding we can say more precisely how the proof goes, and what the issues are that we cannot get into. The argument is in three steps.

Let $f(t)$ be a square integrable function and let $⬚ > 0$.

Step (1): Any function in $L^2([0,1])$ can be approximation in the L^2-norm by a continuously differentiable function. That is, starting with a given f in $L^2([0,1])$ and any $⬚ > 0$ we can find a function $g(t)$ that is continuously differentiable on $[0,1]$ for which

$$\| f - g \| < ⬚.$$

Step (2): From the discussion above, we now know (at least we've now been told, with some indication of why)that the Fourier partial sums for a continuously differentiable function (p $= 1$ in the statement of the theorem) converge uniformly to the function. Thus, with $g(t)$ as in step (1), we can choose N so large that

$max |g(t) - \sum_{n=-N}^{N} g\hat{}(n)e^{2\pi int}| < ⬚.$

Then for the L^2-norm,

$$\int_0^1 \left| g(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right|^2 dt \leq \int_0^1 \left(\max \left| g(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right| \right)^2 dt$$

$$< \int_0^1 \varepsilon^2 dt = \varepsilon^2.$$

Hence

$$\left\| g(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right\| < \square .$$

Step (3): Remember that the Fourier coefficients provide the best finite approximation in L^2 to the function, that is, as we'll need it,

$$\left\| f(t) - \sum_{n=-N}^{N} f\hat{\ }(n)e^{2\pi i n t} \right\| \leq \left\| f(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right\| .$$

And at last

$$\left\| f(t) - \sum_{n=-N}^{N} f\hat{\ }(n)e^{2\pi i n t} \right\| \leq \left\| f(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right\|$$

$$= \left\| f(t) - g(t) + g(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right\|$$

$$\leq \left\| f(t) - g(t) \right\| + \left\| g(t) - \sum_{n=-N}^{N} g\hat{\ }(n)e^{2\pi i n t} \right\| < 2\square .$$

This shows that

$$\left\| f(t) - \sum_{n=-N}^{N} f\hat{\ }(n)e^{2\pi i n t} \right\|$$

Can be made arbitrarily small by taking N large enough.

In the following we study more on the Gibbs phenomenon. Here's what's involved in establishing the Gibbs' phenomenon for the square wave

$$f(t) = \begin{cases} -1 & -1/2 \leq t < 0 \\ +1 & 0 \leq t < +1/2 \end{cases}$$

We're supposed to show that

$$\lim_{N\to\infty} \max S_N(t) = 1.089490\ ....$$

Since we've already introduced the Dirichiet kernel, we'll be content with showing the approach and the outcome, and won't give the somewhat tedious detailed estimates.the partial sum $S_N(t)$ can be written as a convolution with D_N. in the case of the square wave, as we've set up here,

$$S_N(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} D_N\,(t-s)\,f(s)ds$$

$$= -\int_{-\frac{1}{2}}^{0} D_N\,(t-s)ds + \int_{0}^{\frac{1}{2}} D_N\,(t-s)\,ds$$

$$= -\int_{-\frac{1}{2}}^{0} D_N\,(s-t)ds + \int_{0}^{\frac{1}{2}} D_N\,(s-t)ds \quad \text{(using that } D_N \text{is even.)}$$

The idea next is to try to isolate, and estimate, the behavior near the origin by getting an integral from $-t\ to\ t$. we can do this by first making a change of variable $u = s - t$in both integrals. This results in

$$-\int_{-\frac{1}{2}}^{0} D_N\,(s-t)ds + \int_{0}^{\frac{1}{2}} D_N\,(s-t)ds = -\int_{-\frac{1}{2}-t}^{-t} D_N\,(u)du + \int_{-t}^{\frac{1}{2}-t} D_N(u)du\,.$$

To this last expression add and subtract

$$\int_{-t}^{t} D_N\,(u)du$$

And combine integrals to further obtain

$$-\int_{-\frac{1}{2}-t}^{-t} D_N\,(u)du + \int_{-t}^{\frac{1}{2}-t} D_N(u)du$$

$$= -\int_{-\frac{1}{2}-t}^{t} D_N(u)\,du + \int_{-t}^{\frac{1}{2}-t} D_N(u)\,du + \int_{-t}^{t} D_N(u)\,du$$

Finally, make a change of variable $w = -u$ in the first integral and use the evenness of D_N. Then the first two integrals combine and we are left with, again letting s be the variable of integration in both integrals,

$$S_N(t) = \int_{-t}^{t} D_N(s)\,ds - \int_{\frac{1}{2}-t}^{\frac{1}{2}+t} D_N(s)\,ds.$$

The reason that this is helpful is that using the explicit formula for D_N one can show (this takes some work - integration by parts) that

$$\left| S_N(t) - \int_{-t}^{t} D_N(s)\,ds \right| = \left| \int_{\frac{1}{2}-t}^{\frac{1}{2}+t} D_N(s)\,ds \right| \leq \frac{constant}{n},$$

And hence

$$\lim_{N\to\infty} \left| S_N(t) - \int_{-t}^{t} D_N(s)\,ds \right| = 0.$$

This mean that if we can establish for $\int_{-t}^{t} D_N(s)\,ds$ we'll also get one for $S_N(t)$. that, too, takes some work, but the fact that has an explicit formula for D_N makes it possible to deduce for $|t|$ small and N large that $\int_{-t}^{t} D_N(t)\,dt$, and hence $S_N(t)$ is well approximated by

$$\frac{2}{\pi} \int_{0}^{(2N+1)\pi t} \frac{sins}{s}\,ds.$$

This integral has a maximum at the first place where

$\sin\big((2N+1)\pi t\big) = 0$, i.e., at $t = 1/(2N+1)$ at this point the value of the integral (found via numerical approximations) is

$$\frac{2}{\pi} \int_{0}^{\pi} \frac{sins}{s}\,ds = 1.09940\ldots..$$

Chapter (2)

Properties of Convolution and Central Limit Theorem

Section (2.1): $A *$ and Properties of Convolution

We begin this section by studying the $A *$ is Born. Some of the properties of the Fourier transform that we have already derived can be thought of as addressing the question how can we use one signal to modify another? The easiest is the result on additivity according to which

$$\mathcal{F}(f + g) = \mathcal{F}f + \mathcal{F}g$$

Adding the signal $g(t)$ to the signal $f(t)$ adds the amounts $\mathcal{F}g(s)$ to the frequency components $\mathcal{F}f(s)$. (Symmetrically, $f(t)$ modifies $g(t)$ in the same way.) The spectrum of $f + g$ may be more or less "complicated" then the spectrum of f and g alone, and it's on elementary operation in both the time domain and the frequency domain that produces or eliminates the complications. It's on operation that is also easily undone.

We can view the question of using one signal to modify another in either the time domain or in the frequency domain, sometimes with equal ease and sometimes with one point of preferred. We just looked at soms, what about products? The trivial case is multiplying by a constant, as in $\mathcal{F}(af)(s) = a\mathcal{F}f(s)$. Theenergies of the harmonics are all affected by the same amount, so, thinking of music for example, the signalsounds the same, only louder or softer. It's much less obvious how to scale the harmonics separately. Thatis, as a question "in the frequency domain" we ask:

> Is there some combination of the signals $f(t)$ and $g(t)$ so that in the frequency domain theFourier transform is
>
> $$\mathcal{F}g(s)\mathcal{F}f(s) \, ?$$
>
> In other words, in the time domain can we combine the signal $g(t)$ with the signal $f(t)$ so thatthe frequency components $\mathcal{F}f(s)$ of $f(t)$ are scaled by the frequency components $\mathcal{F}g(s)$ of $g(t)$?(Once again this is symmetric — we could say that the frequency components $\mathcal{F}g(s)$ are scaledby the frequency components $\mathcal{F}f(s)$.

The product of the Fourier transforms of $f(t)$ and $g(t)$ is

$$\mathcal{F}g(s)\mathcal{F}f(s) = \int_{-\infty}^{\infty} e^{-2\pi i st}\,d(t)dt \int_{-\infty}^{\infty} e^{-2\pi i sx}\,f(x)dx.$$

We used different variables of integration in the two integrals becausewe're going to combine the productinto an iterated integral.

$$\int_{-\infty}^{\infty} e^{-2\pi i st}\,g(t)dt \int_{-\infty}^{\infty} e^{-2\pi i sx}\,f(x)dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi i st}e^{-2\pi i st}\,g(t)f(x)dt\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi i s(t+x)}\,g(t)f(x)dt\,dx$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-2\pi i s(t+x)}\,g(t)dt\right)f(x)dx$$

Now make the change of variable $u = t + x$in the inner integral. Then$t = u - x$, $du = dt$, and the limitsare the same. The result is

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-2\pi i s(t+x)}\,g(t)dt\right)f(x)dx$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-2\pi i su}\,g(u-x)du\right)f(x)dx$$

Next, switch the order of integration:

$$\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-2\pi i su}\,g(u-x)du\right)f(x)dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi i su}\,g(u-x)\,f(x)du\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi i su}\,g(u-x)\,f(x)dx\,du$$

$$= \int_{-\infty}^{\infty} e^{-2\pi i su}\left(\int_{-\infty}^{\infty} g(u-x)f(x)dx\right)du$$

The inner integral is a function of u. Let's set it up on its own:

$$h(u) = \int_{-\infty}^{\infty} g(u-x)f(x)dx\,.$$

The outer integral produces the Fourier transform of h:

$$\int_{-\infty}^{\infty} e^{-2\pi i su}\left(\int_{-\infty}^{\infty} g(u-x)f(x)dx\right)du = \int_{-\infty}^{\infty} e^{-2\pi i su}h(u)du = \mathcal{F}h(s)$$

Switching the variable name for h from $h(u)$ to $h(t)$ (solely for psychological comfort), we have discovered that the signals $f(t)$ and $g(t)$ are combined into a signal

$$h(t) = \int_{-\infty}^{\infty} g(t-x)f(x)dx.$$

In other words,

$$\mathcal{F}h(s) = \mathcal{F}g(s)\mathcal{F}f(s).$$

We have solved our problem. The only thing to do is to realize what we've done and declare it to the world. We make the following definition:

Definition (2.1.1): (Convolution defined)

The convolution of two functions $g(t)$ and $f(t)$ is the function

$$h(t) = \int_{-\infty}^{\infty} g(t-x)f(x)dx.$$

We use the notation

$$(g * f)(t) = \int_{-\infty}^{\infty} g(t-x)f(x)dx.$$

We can now proudly announce:

Theorem (2.1.2): Convolution Theorem

$$\mathcal{F}(g * f)(s) = \mathcal{F}g(s)\mathcal{F}f(s)$$

In other notation: If $f(t) \leftrightharpoons F(s)$ and $g(t) \leftrightharpoons G(s)$ then
$(g * f)(t) \leftrightharpoons G(s)F(s)$.

In words: Convolution in the time domain corresponds to multiplication in the frequency domain.

Recall that when we studied Fourier series, convolution came up in the form

$$(g * f)(t) = \int_{0}^{1} g(t-x)f(x)dx.$$

In that setting, for the integral to make sense, i.e., to be able to evaluate $g(t-x)$ at points outside the interval from 0 to 1, we had to assume that g was periodic. That's not an issue in the present setting, where we assume that $f(t)$ and $g(t)$ are defined for all t, so the factors in the integral

$$\int_{-\infty}^{\infty} g(t-x)f(x)dx$$

are defined everywhere.

Remark (2.1.3): (Remark on notation, again)

It's common to see the people write the convolution as $g(t) * f(t)$, putting the

variable t in each of g and f. There are times when that's OK, even sometimes preferable to introducing a lot of extra notation, but in general I think it's a bad idea because it can lead to all sorts of abuses and possible mistakes. For example, what's $g(2t) * f(t)$? If you plugged in too casually you might write this as the integral

$$\int_{-\infty}^{\infty} g(2t - x)f(x)dx .$$

That's wrong. The right answer in convolving $g(2t)$ and $f(t)$ is

$$\int_{-\infty}^{\infty} g(2(t - x))f(x)dx = \int_{-\infty}^{\infty} g(2t - 2x)f(x)dx .$$

Now we discuss convolving in the frequency domain. If you look at the argument for the convolution theorem $\mathcal{F}(g * f) = \mathcal{F}g \cdot \mathcal{F}f$, you'll see that we could have carried the whole thing out for the inverse Fourier transform, and given the symmetry between the Fourier transform and its inverse that's not surprising.

That is, we also have

$$\mathcal{F}^{-1}(g * f) = \mathcal{F}^{-1}g \cdot \mathcal{F}^{-1}f .$$

What's more interesting, and doesn't follow without a little additional argument, is this:

$$\mathcal{F}(gf)(s) = (\mathcal{F}g * \mathcal{F}f)(s).$$

In words: Multiplication in the time domain corresponds to convolution in the frequency domain.

Here's how the derivation goes. We'll need one of the duality formulas, the one that says

$$\mathcal{F}(\mathcal{F}f)(s) = f(-s) \text{ or } \mathcal{F}(\mathcal{F}f) = f^- \text{without the variable.}$$

To derive the identity $\mathcal{F}(gf) = \mathcal{F}g * \mathcal{F}f$, we write, for convenience, $h = \mathcal{F}f$ and $k = \mathcal{F}g$. Then we're to show

$$\mathcal{F}(gf) = k * h .$$

The one thing we know is how to take the Fourier transform of a convolution, so, in the present notation, $\mathcal{F}(k * h) = (\mathcal{F}k)(\mathcal{F}h)$. But now $\mathcal{F}k = \mathcal{F}\mathcal{F}g = g^-$, from the identity above, and likewise $\mathcal{F}h = \mathcal{F}\mathcal{F}f = f^-$. So $\mathcal{F}(k * h) = g^- f^- = (gf)^-$, or

$$gf = \mathcal{F}(k * h)^- .$$

Now, finally, take the Fourier transform of both sides of this last equation and appeal to the $\mathcal{F}\mathcal{F}$ identity again:

$$\mathcal{F}(gf) = \mathcal{F}(\mathcal{F}(k * h)^-) = k * h = \mathcal{F}g * \mathcal{F}f .$$

We're done.

Remark (2.1.4):

You may wonder why we didn't start by trying to prove $\mathcal{F}(gf)(s) = (\mathcal{F}g * \mathcal{F}f)(s)$ rather than $\mathcal{F}(g * f) = (\mathcal{F}f)(\mathcal{F}g)$ as we did. That is, it seems more "natural" to multiply signals in the time domainand see what effect this has in the frequency domain, so why not work with $\mathcal{F}(fg)$ directly? But write theintegral for $\mathcal{F}(gf)$,there is nothing you can do with it to get normal $\mathcal{F}g * \mathcal{F}f$.

In the following we illustrate what is Convolution, Really? There's not a single answer to that question. Those of you who have had a course in "Signals and Systems"probably saw convolution in connection with Linear Time Invariant Systems and the impulse response forsuch a system. That's a verynatural setting for convolution.

The fact is that convolution is used in many ways and for many reasons, and it can be a mistake to try toattach to it one particular meaning or interpretation. This multitude of interpretations and applicationsis somewhat like the situation with the definite integral. When you learned about the integral, when you learn about the integral, chancesare that it was introduced via an important motivating problem, typically recovering the distance traveledfrom the velocity, or finding the area under a curve. That's fine, but the integral is really a much moregeneral and flexible concept than those two sample problems might suggest. You do yourself no service ifevery time you think to use an integral you think only of one of those problems. Likewise, you do yourselfno service if you insist on one particular interpretation of convolution.

To pursue the analogy with the integral a little bit further, in pretty much all applications of the integralthere is a general method at work: cut the problem into small pieces where it can be solved approximately,sum up the solution for the pieces, and pass to a limit. There is also often a general method to working, orseeking to work with convolutions: usually there's something that has to do with smoothing and averaging,understood broadly.

For example, in using Fourier series to solve the heat equation on a circle, we saw that the solutionwas expressed as a convolution of the initial heat distribution with the Green's function (or fundamentalsolution). That's a smoothing and averaging interpretation (both) of the convolution. It's also a linearsystems interpretation of convolution, where the system is described by the heat equation.

In brief, we'll get to know the convolution by seeing it in action:
- Convolution is what convolution does.

We now discuss but can I visualize convolution? or "Flip this, buddy".
Again for those of you who have seen convolution in earlier courses,you've probably heard the expression "flip and drag". For

$$(g * f)(t) = \int_{-\infty}^{\infty} g(t - x)f(x)dx$$

here's what this means.

1. Fix a value t. The graph of the function $g(x - t)$ has the same shape as $g(x)$ but shifted to the right by t. Then forming $g(t - x)$ flips the graph (left-right) about the line $x = t$. If the most interesting or important features of $g(x)$ are near $x = 0$, e.g., if it's sharply peaked there, then those featuresare shifted to $x = t$for the function $g(t - x)$ (but there's the extra "flip" to keep in mind).

2. Multiply the two functions $f(x)$ and $g(t - x)$ and integrate with respect to x. Remember that the value of the convolution$(g * f)(t)$ is not just the product of the values of $f$and the flipped and shifted g, it's the integral of the product — much harder to visualize. Integrating the product sums up thesevalues, that's the "dragging" part.

In the following we study Smoothing and averaging. We prefer to think of the convolution operation as using one function tosmooth and average the other. (Say $g$is used to smooth $f$in $g * f$.) In many common applications $g(x)$is a positive function, concentrated near 0, with total area 1,

$$\int_{-\infty}^{\infty} g(x)dx = 1,$$

like a sharply peaked Gaussian, for example (stay tuned). Then $g(t - x)$ is concentrated near $t$and stillhas area 1. For a fixed t, forming the integral

$$\int_{-\infty}^{\infty} g(t - x)f(x)dx$$

is like taking a weighted average of the values of $f(x)$ near $x = t$, weighted by the values of (the flippedand shifted) g. (It's a legitimate weighted average because$\int_{-\infty}^{\infty} g(x)dx = 1$.)

That's the averaging part of the description: Computing the convolution $g * f$at $t$replaces the value $f(t)$by a weighted average of the values of $f$near t. Where does the smoothing come in? Here's where.

Changing t("dragging" $g(t - x)$ through different values of t) repeats this operation.

Again take the case of an averaging-type function $g(t)$, as above. At a given value of t, $(g * f)(t)$ is aweighted average of values of f near t. Move ta little to a point t'. Then $(g * f)(t')$ is a weighted averageof values of f near t', which will include values of $f$that entered into the average near t. Thus the valuesof the convolutions $(g * f)(t)$ and $(g * f)(t')$ will likely be closer to each other than are the values $f(t)$ and

$f(t')$. That is, $(g * f)(t)$ is "smoothing" $f$as $t$varies — there's less of a change between values of theconvolution than between values of f.

We see example of smoothing. The rectfunction $\Pi(x)$ is discontinuous — it has jumps at $\pm 1/2$. The convolution $\Pi * \Pi$ is the triangle functionΛ, which is continuous — the jumps at the endpoints have been smoothed out. There's still a corner, butthere's no discontinuity.

In fact, as an aphorism we can state

- The convolution $g * f$is at least as smooth a function as $g$and $f$are separately.

In the following we discuss A smear job, too Now, be a little careful in how you think about this averaging and smoothing process.Computing any value of $(g * f)(t)$ involves all of the values of $g$and all of the values of f, and adding theproducts of corresponding values of $g$and $f$with one of the functions flipped and dragged. If both $f(t)$ and$g(t)$ become identically zero after awhile then the convolution $g * f$will also be identically zero outside ofsome interval. But if either $f(t)$ or $g(t)$ does not become identically zero then neither will the convolution.

In addition to averaging and smoothing the convolution also "smears" out the factors — not a becomingdescription, but an accurate one.

Definitely keep the general description we've just gone through in mind, but as far as visualizing theconvolution of any two old functions, we think it's of dubious value to beat yourself up trying to do that.It's hard geometrically, and it's hard computationally, in the sense that one have to calculate some tediousintegrals.

By the way, of course we can try to get some intuition for how the convolution looks by thinking of what'shappening in the frequency domain. It's not so farfetched to try to imagine the Fourier transforms $\mathcal{F}f$,

$\mathcal{F}g$, and their product, and then imagine the inverse transform to get our$g * f$.

Now we study properties of Convolution: It's a Lot like MultiplicationConvolution behaves in many ways (not all ways) like multiplication. For example, it is commutative:

$$f * g = g * f.$$

So although it looks like the respective roles of f and g are different — one is "flipped and dragged", the other isn't — in fact they share equally in the end result.

We defined the convolution so that the convolution theorem holds, that is so that $\mathcal{F}(g * f) = \mathcal{F}g\mathcal{F}f$. But g and f enter symmetrically on the right hand side, so $g * f = f * g$ — $g(t)$ can be used to modify $f(t)$ or $f(t)$ can be used to modify $g(t)$.

Nevertheless, the commutativity property is easy to check from the definition:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t - u)g(u)du$$

$$= \int_{-\infty}^{\infty} g(t - v)f(v)dv \text{ (making the substitution} v = t - u)$$

$$= (g * f)(t).$$

The same idea, a change of variable but with more bookkeeping, establishes that convolution is associative:

$$(f * g) * h = f * (g * h).$$

Much more easily one gets that

$$f * (g + h) = f * g + f * h.$$

The corresponding statements are easily verified in the frequency domain.

How about a "1"? Is there a function which is to convolution as 1 is to multiplication? Is there a function g such that

$$(g * f)(t) = f(t), \qquad \text{for all functions} f?$$

What property would such a g have? Take Fourier transforms of both sides:

$$\mathcal{F}f(s)\mathcal{F}g(s) = \mathcal{F}f(s).$$

Then $g(x)$ must be such that

$$\mathcal{F}g(s) = 1.$$

Is there such a g? Applying the inverse Fourier transform would lead to

$$\int_{-\infty}^{\infty} e^{2\pi i s x}dx,$$

and that integral does not exist — even we wouldn't try to slip that by the rigor police. Something is up here. Maybe Fourier inversion doesn't work in this case, or else there's no classical function whose Fourier transform is 1, or something. In fact, though the integral does not exist in any sense, the problem of a "1 for convolution" leads exactly to the delta function, or unit impulse — not a classical function, but a "generalized" function.

How about "division"? Suppose we know h and g in

$$h = f * g$$

and we want to solve for f. Again, taking Fourier transforms we would say

$$\mathcal{F}h = \mathcal{F}f \cdot \mathcal{F}g \Rightarrow \mathcal{F}f = \frac{\mathcal{F}h}{\mathcal{F}g}.$$

We'd like the convolution quotient to be the inverse Fourier transform of $\mathcal{F}h/\mathcal{F}g$. But there are problemscaused by places where $\mathcal{F}g = 0$, along with the usual problems with the integral for the inverse Fouriertransform to exist.

Solving for $f(t)$ is the deconvolution problem, which is extremely important in applications. Many timesa noisy signal comes to you in the form $h = f * g$; the signal is f, the noise is g, you receive h. You makesome assumptions about the nature of the noise.

Now we study the identities. It's not hard to combine the various rules we have and develop an algebra of convolutions. Such identities can be of great use — it beats calculating integrals. Here's an assortment. (Lowerand uppercase letters are Fourier pairs.)

$$\big((f \cdot g) * (h \cdot k)\big)(t) \leftrightharpoons \big((F * G) \cdot (H * K)\big)(s)$$

$$\big((f(t) + g(t)) \cdot (h(t) + k(t))\big) \leftrightharpoons \Big(\big((F + G) * (H + K)\big)\Big)(s)$$

$$(f(t) \cdot (g * h)(t) \leftrightharpoons \big(F * (G \cdot H)\big)(s)$$

Section (2.2): Convolution in Action and Central Limit Theory

We begin this section by studying the convolution in action. I: A Little Bit on Filtering. "Filtering" is a generic term for just about any operation one might want to apply to a signal. We haveto be reasonable, of course — there's usually some feature of the signal that one wants to enhance oreliminate, and one expects something of the original signal to be recognizable or recoverable after it's been filtered. Most filters are described as somehow modifying the spectral content of a signal, and they arethus set up as an operation on the Fourier transform of a signal. But it's worthwhile saying a little bit now because themost common filters operate through multiplication in the frequency domain, hence through convolutionin the time domain.

The features are:

- An input signal $v(t)$
- An output signal $w(t)$
- The operation that produces $w(t)$ from $v(t)$ in the time domain is convolution with a function $h(t)$:

$$w(t) = (h * v)(t) = \int_{-\infty}^{\infty} h(t - x)v(x)dx$$

With this description the Fourier transforms of the input and output are related by multiplication in thefrequency domain:
$$W(s) = H(s)V(s),$$
where, following tradition, we denote the Fourier transforms by the corresponding capital letters. In thiscontext $h(t)$ is usually called the impulse response and $H(s)$ is called the transfer function. It seems tobe a matter of course always to denote the impulse response by $h(t)$ and always to denote the transferfunction by $H(s)$.

Remember that $h(t)$, hence $H(s)$, is "fixed" in this discussion. It's wired into the circuit or coded into thesoftware and it does what it does to any input you may give it. Filters based on convolution are usuallydesigned to have a specific effect on the spectrum of an input, and so to design a filter is to design a transferfunction. The operations, which you're invited to draw a block diagram for, are thus

Input $\to$ Fourier transform $\to$ Multiply by H $\to$ Inverse Fourier transform = output

We want to see some examples of this— filters that are in day-to-day use and the principles thatgo into their design.

One preliminary comment about how the spectra of the input and output are related. Write
$$V(s) = |V(s)|e^{i\phi_V(s)}, \qquad \phi_V(s) = \tan^{-1}\left(\frac{\operatorname{Im} V(s)}{\operatorname{Re} V(s)}\right),$$
so the phase of $V(s)$ is $\phi_V(s)$, with similar notations for the phases of $W(s)$ and $H(s)$. Then
$$|W(s)|e^{i\phi_W(s)} = |H(s)|e^{i\phi_H(s)}|V(s)|e^{i\phi_V(s)} = |H(s)||V(s)|e^{i(\phi_H(s)+\phi_V(s))}.$$
Thus the magnitudes multiply and the phases add:
$$|W(s)| = |H(s)||V(s)|$$
$$\phi_W(s) = \phi_V(s) + \phi_H(s)$$
Multiplying $V(s)$ by $H(s)$ can't make the spectrum of $V(s)$ any bigger, but it can make the spectrumsmaller by zeroing out parts of it. Furthermore, there is no phase change when $\phi_H(s) = 0$, and thishappens when $H(s)$ is real. In this case only the amplitude is changed when the signal goes through thefilter. Common examples of filters that do both of these things — modify some part of the magnitude ofthe spectrum with no phase change — are lowpass, bandpass, highpass, and notch filters, to which we'llnow turn.

In the following we discuss Designing filters, Lowpass filters An ideal lowpass filter cuts off all frequencies above a certain amount v_c ("c" for

"cutoff")and lets all frequencies below v_c pass through unchanged. (Hence the description "lowpass".) If we writethe operation as

$$w(t) = (h * v)(t) \leftrightharpoons W(s) = H(s)V(s),$$

then the transfer function we want is

$$H(s) = \begin{cases} 1 & |s| < v_c \\ 0 & |s| \geq v_c \end{cases}$$

Multiplying $V(s)$ by $H(s)$ leaves unchanged the spectrum of $v$for $|s| < v_c$ and eliminates the otherfrequencies. The transfer function is just a scaled rect function, and we can write it (to remind you) as

$$H(s) = \Pi_{2v_c}(s) = \Pi\left(\frac{s}{2v_c}\right) = \begin{cases} 1 & \left|\frac{s}{2v_c}\right| < \frac{1}{2} \\ 0 & \left|\frac{s}{2v_c}\right| \geq \frac{1}{2} \end{cases} = \begin{cases} 1 & |s| < v_c \\ 0 & |s| \geq v_c \end{cases}$$

In the time domain the impulse response is the inverse Fourier transform of Π_{2v_c}, and this is

$$h(t) = 2v_c \text{sinc}(2v_c t).$$

By the way, why is this called just a "lowpass filter"; aren't the frequencies below $-v_c$ also eliminatedand so not "passed" by the filter? Yes, but remember that for real signals $v(t)$ (which is where this isapplied) one has the symmetry relation $V(-s) = \overline{V(s)}$. The positive and negative frequencies combine inreconstructing the real signal in the inverse Fourier transform, much like what happens with Fourier series.

Thus one wants to pass the frequencies with $-v_c < s < v_c$ and eliminate the frequencies with $s \geq v_c$ and$s \leq -v_c$.

And, by the way, why is this called an ideal lowpass filter? Because the cutoff is a sharp one — right ata particular frequency v_c. In practice this cannot be achieved, and much of the original art of filter designis concerned with useful approximations to a sharp cutoff.

Now we illustrate Bandpass filters. Another very common filter passes a particular band of frequencies through unchangedand eliminates all others. This is the ideal bandpass filter. Its transfer function, $B(s)$, can be constructedby shifting and combining the transfer function $H(s)$ for the lowpass filter.

We center our bandpass filter at $\pm v_0$ and cut off frequencies more than v_c above and below v_0; just as forthe lowpass filter we pass symmetric bands of positive frequencies and negative frequencies, and eliminate everything else. That is we define the transfer function of a bandpass filter to be

$$B(s) = \begin{cases} 1 & v_0 - v_c < |s| < v_0 + v_c \\ 0 & otherwise \end{cases}$$
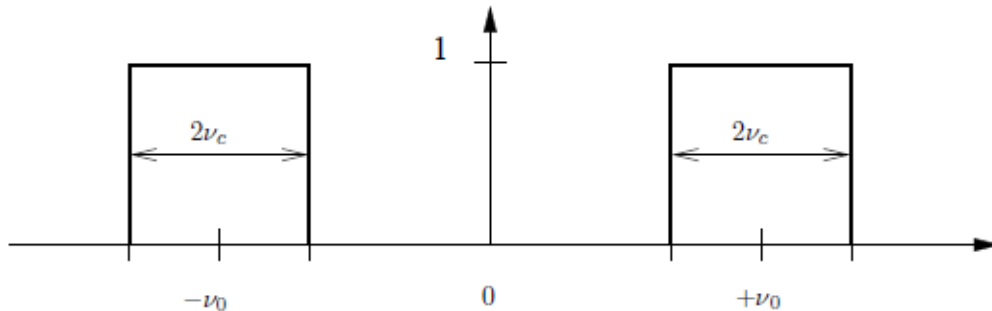$$= H(s - v_0) + H(s + v_0)$$

Here's the graph.

Figure (2.1)

From the representation of $B(s)$ in terms of $H(s)$ it's easy to find the impulse response, $b(t)$. That's givenby

$$b(t) = h(t)e^{2\pi i v_0 t} + h(t)e^{-2\pi i v_0 t}$$

(using the shift theorem or the modulation theorem)

$$= 4v_c \cos(2\pi v_0 t)\operatorname{sinc}(2v_c t).$$

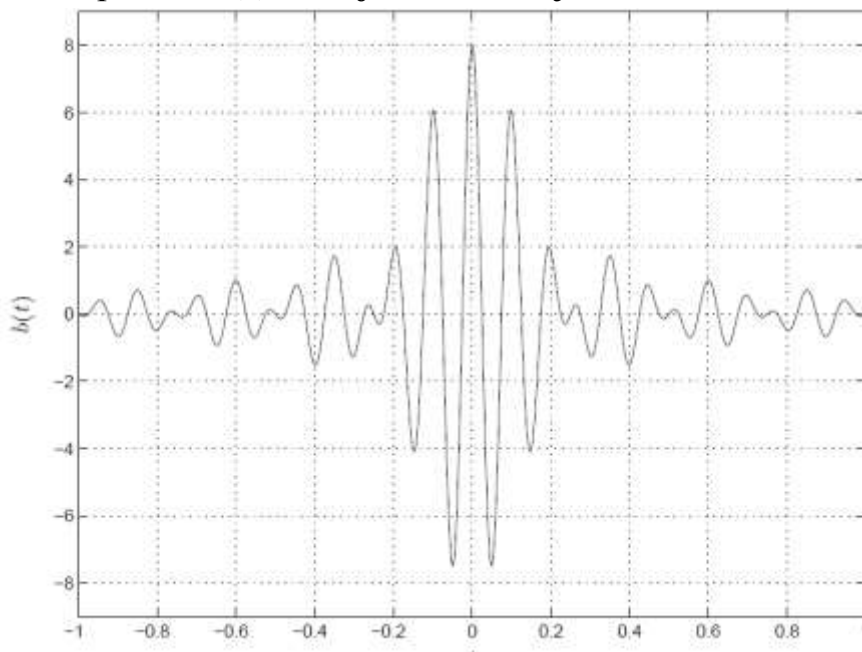Here's a plot of $b(t)$ for $v_0 = 10$ and $v_c = 2$:

Figure (2.2)

Now we study highpass filters. The twin to an ideal lowpass filter is an ideal high pass filter, where all frequenciesabove a cutoff frequency v_c are passed through unchanged and all frequencies below are eliminated. Youmight use this, for example, if there's a slow "drift" in your data that suggests a low

frequency disturbanceor noise that you may want to eliminate. Highpass filters are used on images to sharpen edges and details(associated with high spatial frequencies) and eliminate blurring (associated with low spatial frequencies). The graph of the transfer function for an ideal highpass filter looks like:
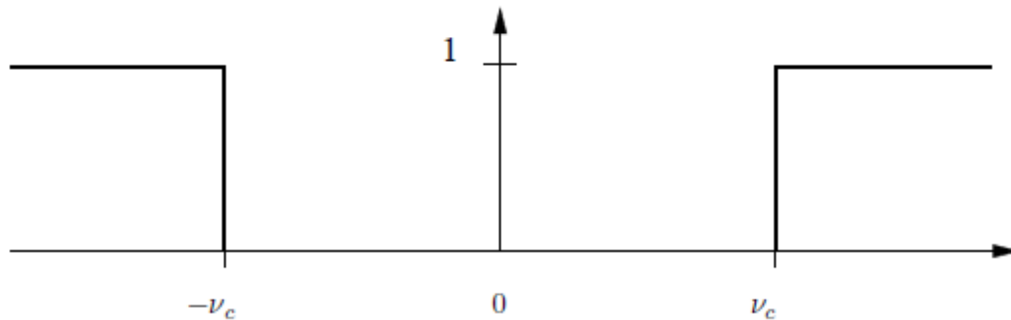

Figure (2.3)

It's easy to write a formula for this; it's just

$$\text{High}(s) = 1 - \Pi_{2v_c}(s),$$

where v_c is the cutoff frequency. At this point we're stuck. We can't find the impulse response becausewe haven't yet gained the knowledge that the inverse Fourier transform of 1 is the δ function. Think ofthe highpass filter as the evil twin of the lowpass filter.

In the following we discuss Notch filters. The evil twin of a bandpass filter is a notch filter. The effect of a notch filter is to eliminatefrequencies within a given band (the "notch") and to pass frequencies outside that band. To get the transferfunction we just subtract a bandpass transfer function from 1. Using the one we already have:

$$\text{Notch}(s) = 1 - B(s) = 1 - \big(H(s - v_0) + H(s + v_0)\big).$$

This will eliminate the positive frequencies between $v_0 - v_c$ and $v_0 + v_c$, and pass all frequencies outside of these two bands.
You can draw your own graph of that.

Unfortunately, for the impulse response we're in the same position here as we were for the highpass filter.We cannot write down the impulse response without recourse to δ's, so this will have to wait.

Now we study Convolution in Action II: Differential Equations
One of the most common uses of convolution and the Fourier transform is in solving differential equations.Solving differential equations was Fourier's original motivation for Fourier series and the use of the Fouriertransform to this end has continued to exercise a strong influence on the theory and the applications. We'llconsider several illustrations, from a simple ordinary

differential equation to problems associated with the heat equation. We'll also revisit the problem of a signal propagating along a cable.

The derivative formula To put the Fourier transform to work, we need a formula for the Fourier transform of the derivative, and as you found in homework:

$$(\mathcal{F}f')(s) = 2\pi i s \mathcal{F}f(s).$$

We see that differentiation has been transformed into multiplication, another remarkable feature of the Fourier transform and another reason for its usefulness. Formulas for higher derivatives also hold, and the result is:

$$\left(\mathcal{F}f^{(n)}\right)(s) = (2\pi i s)^n \mathcal{F}f(s).$$

In general, a differential operator can be thought of as a polynomial in d/dx, say of the form

$$P\left(\frac{d}{dx}\right) = a_n \left(\frac{d}{dx}\right)^n + a_{n-1}\left(\frac{d}{dx}\right)^{n-1} + \cdots + a_1 \frac{d}{dx} + a_0,$$

and when applied to a function $f(x)$ the result is

$$a_n f^{(n)} + a_{n-1}f^{(n-1)} + \cdots + a_1 f' + a_0 f.$$

If we now take the Fourier transform of this expression, we wind up with the Fourier transform of f multiplied by the corresponding n-th degree polynomial evaluated at $2\pi i s$.

$$\left(\mathcal{F}\left(p\left(\frac{d}{dx}\right)f\right)\right)(s) = P(2\pi i s)\mathcal{F}f(s)$$

$$= (a_n(2\pi i s)^n + a_{n-1}(2\pi i s)^{n-1} + \cdots + a_n(2\pi i s) + a_0)\mathcal{F}f(s).$$

Don't underestimate how important this is.

A simple ordinary differential equation and how to solve it. You might like starting off with the classic second order, ordinary differential equation

$$u'' - u = -f$$

Maybe you've looked at a different form of this equation, but I'm writing it this way to make the subsequent calculations a little easier. $f(t)$ is a given function and you want to find $u(t)$.

Take the Fourier transform of both sides:

$$(2\pi i s)^2 \mathcal{F}u - \mathcal{F}u = -\mathcal{F}f$$
$$-4\pi^2 s^2 \mathcal{F}u - \mathcal{F}u = -\mathcal{F}f$$
$$(1 + 4\pi^2 s^2)\mathcal{F}u = \mathcal{F}f$$

So we can solve for $\mathcal{F}u$ as

$$\mathcal{F}u = \frac{1}{1 + 4\pi^2 s^2}\mathcal{F}f,$$

and — with a little struggle — we recognize $\frac{1}{(1+4\pi^2 s^2)}$ as the Fourier transform of $\frac{1}{2}e^{-|t|}$, that is,

$$\mathcal{F}u = \mathcal{F}\left(\frac{1}{2}e^{-|t|}\right).\mathcal{F}f.$$

The right hand side is the product of two Fourier transforms. Therefore, according to the convolutiontheorem,

$$u(t) = \frac{1}{2}e^{-|t|} * f(t).$$

Written out in full this is

$$u(t) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-|t-\tau|}f(\tau)d\tau.$$

And there you have the two-sided exponential decay in action, as well as convolution.

The heat equation Remember the heat equation? In one spatial dimension, the equation that describesthe rates of change of the temperature $u(x,t)$ of the body at a point $x$and time t(with some normalizationof the constants associated with the material) is the partial differential equation

$$u_t = \frac{1}{2}u_{xx}.$$

In our earlier work on Fourier series we considered heat flow on a circle, and we looked for solutions thatare periodic function of $x$on the interval [0, 1], so u was to satisfy $u(x+1,t) = u(x,t)$. This time we wantto consider the problem of heat flow on the "infinite rod". A rod of great length (effectively of infinitelength) is provided with an initial temperature distribution $f(x)$ and we want to find a solution $u(x,t)$ ofthe heat equation with

$$u(x,0) = f(x).$$

Both $f(x)$ and $u(x,t)$ are defined for $-\infty < x < \infty$, and there is no assumption of periodicity. Knowingthe Fourier transform of the Gaussian is essential for the treatment we're about to give.The idea is to take the Fourier transform of both sides of the heat equation, "with respect to x". TheFourier transform of the right hand side of the equation, $\frac{1}{2}u_{xx}(x,t)$, is

$$\frac{1}{2}\mathcal{F}u_{xx}(s,t) = \frac{1}{2}(2\pi is)^2\mathcal{F}u(s,t) = -2\pi^2 s^2\mathcal{F}u(s,t),$$

from the derivative formula. Observe that the "frequency variable" $s$is now in the first slot of the trans-formed function and that the time variable t is just

going along for the ride. For the left hand side, $u_t(x,t)$,we do something different. We have

$$\mathcal{F}u_t(s,t) = \int_{-\infty}^{\infty} u_t(x,t)e^{-2\pi isx}\,dx \text{ (Fourier transform in } x)$$

$$= \int_{-\infty}^{\infty} \frac{\partial}{\partial t}u(x,t)e^{-2\pi isx}\,dx = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} u(x,t)e^{-2\pi isx}\,dx$$

$$= \frac{\partial}{\partial t}\hat{u}(s,t).$$

Thus taking the Fourier transform (with respect to x) of both sides of the equation

$$u_t = \frac{1}{2}u_{xx}$$

leads to

$$\frac{\partial \mathcal{F}u(s,t)}{\partial t} = 2\pi^2 s^2 t\, \mathcal{F}u(s,t).$$

This is a differential equation in t— an ordinary differential equation, despite the partial derivative symbol— and we can solve it:

$$\mathcal{F}u(s,0) = \mathcal{F}u(s,0)e^{-2\pi^2 s^2 t}.$$

What is the initial condition, $\mathcal{F}u(s,0)$?

$$\mathcal{F}u(s,0) = \int_{-\infty}^{\infty} u(x,0)e^{-2\pi isx}\,dx = \int_{-\infty}^{\infty} f(x)e^{-2\pi isx}\,dx = \mathcal{F}f(s)$$

Putting it all together,

$$\mathcal{F}u(s,t) = \mathcal{F}f(s)e^{-2\pi^2 s^2 t}.$$

We recognize (we are good) that the exponential factor on the right hand side is the Fourier transform ofthe Gaussian,

$$g(x,t) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}.$$

We then have a product of two Fourier transforms,

$$\mathcal{F}u(s,t) = \mathcal{F}g(s,t)\mathcal{F}f(s)$$

and we invert this to obtain a convolution in the spatial domain:

$$u(x,t) = g(x,t) * f(x) = \left(\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}\right) * f(x) \text{(convolution in } x)$$

or, written out,

$$u(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}}e^{-(x-y)^2/2t}\,f(y)\,dy.$$

It's reasonable to believe that the temperature $u(x, t)$ of the rod at a point x at a time $t > 0$ is some kindof averaged, smoothed version of the initial temperature $f(x) = u(x, 0)$. That's convolution at work.
The function

$$g(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.$$

is called the heat kernel (or Green's function, or fundamental solution) for the heat equation for the infiniterod. Here are plots of $g(x, t)$, as a function of x, for $t = 1, 0.5, 0.1, 0.05, 0.01$.
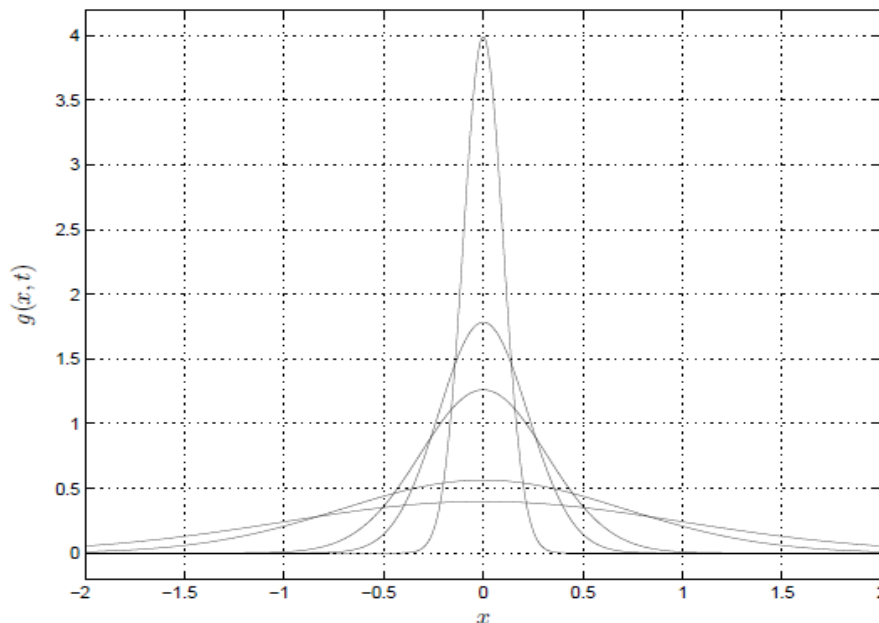

Figure (2.4)

You can see that the curves are becoming more concentrated near $x = 0$. Nevertheless, they are doing soin a way that keeps the area under each curve 1. For

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{\frac{x^2}{2t}} dx = \frac{1}{\sqrt{2\pi t}} e^{-\pi u^2} \sqrt{2\pi t} \, du$$

$$\left(\text{making the substitution} u = \frac{x}{\sqrt{2\pi t}}. \right)$$

$$= \int_{-\infty}^{\infty} e^{-\pi u^2} du = 1$$

More on diffusion — back to the cable Recall from our earlier discussion that William Thomsonappealed to the heat equation to study the delay in a signal sent along a long, undersea telegraph cable. Thephysical intuition, as of the mid 19th century, was that charge "diffused" along the cable. To reconstructpart of

Thomson's solution (essentially) we must begin with a slightly different setup. The equation is thesame

$$u_t = \frac{1}{2} u_{xx},$$

so we're choosing constants as above and not explicitly incorporating physical parameters such as resistanceper length, capacitance per length, etc., but the initial and boundary conditions are different.

We consider a semi-infinite rod, having one end (at $x = 0$) but effectively extending infinitely in thepositive x-direction. Instead of an initial distribution of temperature along the entire rod, we consider asource of heat (or voltage) $f(t)$ at the end $x = 0$. Thus we have the initial condition

$$u(0, t) = f(t).$$

We suppose that

$$u(x, 0) = 0,$$

meaning that at $t = 0$ there's no temperature (or charge) in the rod. We also assume that $u(x, t)$ and itsderivatives tend to zero as $x \to \infty$. Finally, we set

$$u(x, t) = 0 \qquad \text{for} x < 0$$

so that we can regard $u(x, t)$ as defined for all x. We want a solution that expresses $u(x, t)$, the temperature(or voltage) at a position $x > 0$ and time $t > 0$ in terms of the initial temperature (or voltage) $f(t)$ at theendpoint $x = 0$. The analysis of this is really involved.

First take the Fourier transform of $u(x, t)$ with respect to x(the notation $\hat{u}$seems more natural here):

$$\hat{u}(s, t) = \int_{-\infty}^{\infty} e^{-2\pi i s x} u(x, t) dx.$$

Then, using the heat equation,

$$\frac{\partial}{\partial t} \hat{u}(s, t) = \int_{-\infty}^{\infty} e^{-2\pi i s x} \frac{\partial}{\partial t} u(x, t) dx = \int_{-\infty}^{\infty} e^{-2\pi i s x} \frac{\partial^2}{\partial x^2} \frac{1}{2} u(x, t) dx.$$

We need integrate only from 0 to ∞since $u(x, t)$ is identically 0 for $x < 0$. We integrate by parts once:

$$\int_0^{\infty} e^{-2\pi i s x} \frac{\partial^2}{\partial x^2} u(x, t) dx$$

$$= \frac{1}{2} \left(\left[e^{-2\pi i s x} \frac{\partial}{\partial x} u(x, t) \right]_{x=0}^{x=\infty} + 2\pi i s \int_0^{\infty} \frac{\partial}{\partial x} u(x, t) e^{-2\pi i s x} dx \right)$$

$$= -\frac{1}{2} u_x(0, t) + \pi i s \int_0^{\infty} \frac{\partial}{\partial x} u(x, t) e^{-2\pi i s x} dx,$$

taking the boundary conditions on $u(x,t)$ into account. Now integrate by parts a second time:

$$\int_0^\infty \frac{\partial}{\partial x} u(x,t) e^{-2\pi i s x} dx$$

$$= \left[e^{-2\pi i s x} \frac{\partial}{\partial x} u(x,t) \right]_{x=0}^{x=\infty} + 2\pi i s \int_0^\infty e^{-2\pi i s t} u(x,t) dx$$

$$= -u(0,t) + 2\pi i s \int_0^\infty e^{-2\pi i s t} u(x,t) dx$$

$$= -f(t) + 2\pi i s \int_{-\infty}^\infty e^{-2\pi i s t} u(x,t) dx$$

(we drop the bottom limit back to $-\infty$ to bring back the Fourier transform)

$$= -f(t) + 2\pi i s \hat{u}(s,t).$$

Putting these calculations together yields

$$\frac{\partial}{\partial t} \hat{u}(s,t) = -\frac{1}{2} u_x(0,t) - \pi i s \, f(t) - 2\pi^2 s^2 \hat{u}(s,t).$$

Now, this is a linear, first order, ordinary differential equation (in t) for $\hat{u}$. It's of the general type

$$y'(t) + P(t)y(t) = Q(t),$$

and if you cast your mind back and search for knowledge from the dim past you will recall that to solve such an equation you multiply both sides by the integrating factor

$$e^{\int_0^t P(\tau) d\tau}$$

which produces

$$\left(y(t) e^{\int_0^t P(\tau) d\tau} \right)' = e^{\int_0^t P(\tau) d\tau} Q(t).$$

From here you get $y(t)$ by direct integration. For our particular application we have

$$P(t) = 2\pi^2 s^2$$

(that's a constant as far as we're concerned because there's no t)

$$Q(t) = -\frac{1}{2} u_x(0,t) - \pi i s \, f(t).$$

The integrating factor is $e^{2\pi^2 s^2 t}$ and we're to solve

$$\left(e^{2\pi^2 s^2 t} \hat{u}(t) \right)' = e^{2\pi^2 s^2 t} \left(-\frac{1}{2} u_x(0,t) - \pi i s \, f(t) \right).$$

Write τ for t and integrate both sides from 0 to t with respect to :

$$e^{2\pi^2 s^2 t}\hat{u}(s,t) - \hat{u}(s,0) = \int_0^t e^{2\pi^2 s^2 \tau}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)d\tau.$$

But $\hat{u}(s,0) = 0$ since $u(x,0)$ is identically 0, so

$$\hat{u}(s,t) = e^{-2\pi^2 s^2 t}\int_0^t e^{2\pi^2 s^2 \tau}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)d\tau$$

$$= \int_0^t e^{-2\pi^2 s^2 (t-\tau)}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)d\tau.$$

We need to take the inverse transform of this to get $u(x,t)$. Be not afraid:

$$u(x,t) = \int_{-\infty}^{\infty} e^{2\pi isx}\hat{u}(s,t)ds$$

$$= \int_{-\infty}^{\infty} e^{2\pi isx}\left(\int_0^t e^{-2\pi^2 s^2 (t-\tau)}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)d\tau\right)ds$$

$$= \int_0^t \int_{-\infty}^{\infty} e^{2\pi isx} e^{-2\pi^2 s^2 (t-\tau)}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)ds\,d\tau.$$

Appearances to the contrary, this is not hopeless. Let's pull out the inner integral for further examination:

$$\int_{-\infty}^{\infty} e^{2\pi isx}\left(e^{-2\pi^2 s^2 (t-\tau)}\left(-\frac{1}{2}u_x(0,\tau) - \pi is\, f(\tau)\right)\right)ds$$

$$= -\frac{1}{2}u_x(0,\tau)\int_{-\infty}^{\infty} e^{2\pi isx} e^{-2\pi^2 s^2 (t-\tau)}ds$$

$$- \pi is\, f(\tau)\int_{-\infty}^{\infty} e^{2\pi isx} e^{-2\pi^2 s^2 (t-\tau)}ds$$

The first integral is the inverse Fourier transform of a Gaussian; we want to find $\mathcal{F}^{-1}\left(e^{-2\pi s^2 (t-\tau)}\right)$. Recall the formulas

$$\mathcal{F}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right) = e^{-2\pi^2 \sigma^2 s^2}\ ,\mathcal{F}\left(e^{-x^2/2\sigma^2}\right) = \sigma\sqrt{2\pi}e^{-2\pi^2 \sigma^2 s^2}.$$

Apply this with

$$\sigma = \frac{1}{2\pi\sqrt{(t-\tau)}}.$$

Then, using duality and evenness of the Gaussian, we have

$$\int_{-\infty}^{\infty} e^{2\pi isx} e^{-2\pi s^2 (t-\tau)}ds = \mathcal{F}^{-1}\left(e^{-2\pi s^2 (t-\tau)}\right) = \frac{e^{\frac{-x^2}{2(t-\tau)}}}{2\pi\sqrt{(t-\tau)}}.$$

In the second integral we want to find $\mathcal{F}^{-1}\left(s\, e^{-2\pi^2 s^2 (t-\tau)}\right)$. For this, note that

$$s\, e^{-2\pi^2 s^2 (t-\tau)} = \frac{1}{4\pi^2 (t-\tau)} \frac{d}{ds} e^{-2\pi^2 s^2 (t-\tau)}$$

and hence

$$\int_{-\infty}^{\infty} e^{2\pi i s x}\, s\, e^{-2\pi^2 s^2 (t-\tau)} ds = \mathcal{F}^{-1}\left(\frac{1}{4\pi^2 (t-\tau)} \frac{d}{ds} e^{-2\pi^2 s^2 (t-\tau)}\right)$$

$$= \frac{1}{4\pi^2 (t-\tau)} \mathcal{F}^{-1}\left(\frac{d}{ds} e^{-2\pi^2 s^2 (t-\tau)}\right).$$

We know how to take the inverse Fourier transform of a derivative, or rather we know how to take the(forward) Fourier transform, and that's all we need by another application of duality. We use, for a generalfunction f,

$$\mathcal{F}^{-1} f' = (\mathcal{F} f')^- = (2\pi i x \mathcal{F} f)^- = -2\pi i x (\mathcal{F} f)^- = -2\pi i x \mathcal{F}^{-1} f\,.$$

Apply this to

$$\mathcal{F}^{-1}\left(\frac{d}{ds} e^{-2\pi^2 s^2 (t-\tau)}\right) = -2\pi i x \mathcal{F}^{-1}\left(e^{-2\pi^2 s^2 (t-\tau)}\right)$$

$$= -2\pi i x \frac{1}{\sqrt{2\pi(t-\tau)}} e^{-x^2/2(t-\tau)} \text{(from our earlier calculation, fortunately)}$$

Then

$$\frac{1}{4\pi^2 (t-\tau)} \mathcal{F}^{-1}\left(\frac{d}{ds} e^{-2\pi^2 s^2 (t-\tau)}\right) = \frac{2\pi i x}{4\pi^2 (t-\tau)} \frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}}$$

$$= \frac{i}{2\pi} \frac{x\, e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,.$$

That is,

$$\mathcal{F}^{-1}\left(s\, e^{-2\pi^2 s^2 (t-\tau)}\right) = \frac{i}{2\pi} \frac{x\, e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,.$$

Finally getting back to the expression for $u(x,t)$, we can combine what we've calculated for the inverseFourier transforms and write

$$u(x,t) = -\frac{1}{2} \int_0^t u_x(0,\tau)\, \mathcal{F}^{-1}\left(e^{-2\pi s^2 (t-\tau)}\right) d\tau$$

$$- \pi i \int_0^t f(\tau)\, \mathcal{F}^{-1}\left(s\, e^{-2\pi^2 s^2 (t-\tau)}\right) d\tau$$

$$= -\frac{1}{2} \int_0^t u_x(0,\tau) \frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}} d\tau + \frac{1}{2} \int_0^t f(\tau) \frac{x\, e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\, d\tau.$$

We're almost there. We'd like to eliminate $u_x(0,\tau)$ from this formula and express $u(x,t)$ in terms of $f(t)$only. This can be accomplished by a very clever, and We'd say highly nonobvious observation. We knowthat $u(x,t)$ is

zero for $x < 0$; we have defined it to be so. Hence the integral expression for $u(x, t)$ is zerofor $x < 0$. Because of the evenness and oddness in $x$of the two integrands this has a consequence for thevalues of the integrals when $x$is positive. (The first integrand is even in $x$and the second is odd in x.) Infact, the integrals are equal!

Let us explain what happens in a general situation, stripped down, so we can see the idea. Suppose wehave

$$\Phi(x, t) = \int_0^t \phi(x, \tau)\, d\tau + \int_0^t \psi(x, \tau)\, d\tau$$

where we know that: $\Phi(x, t)$ is zero for $x < 0$; $\phi(x, \tau)$ is even in x; $\psi(x, \tau)$ is odd in x. Take $a > 0$. Then$\Phi(-a, \tau) = 0$, hence using the evenness of $\phi(x, \tau)$ and the oddness of $\psi(x, \tau)$,

$$0 = \int_0^t \phi(-a, \tau)\, d\tau + \int_0^t \psi(-a, \tau)\, d\tau = \int_0^t \phi(a, \tau)\, d\tau - \int_0^t \psi(a, \tau)\, d\tau.$$

We conclude that for all $a > 0$,

$$\int_0^t \phi(a, \tau)\, d\tau = \int_0^t \psi(a, \tau)\, d\tau,$$

and hence for $x > 0$ (writing $x$for a)

$$\Phi(x, t) = \int_0^t \phi(x, \tau)\, d\tau + \int_0^t \psi(x, \tau)\, d\tau = 2\int_0^t \psi(x, \tau)\, d\tau$$

$$= 2\int_0^t \phi(x, \tau)\, d\tau \qquad \text{(either } \phi \text{ or } \psi \text{ could be used).}$$

We apply this in our situation with

$$\phi(x, \tau) = -\frac{1}{2}u_x(0, \tau)\frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}}, \psi(x, \tau) = \frac{1}{2}f(\tau)\frac{x\, e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}.$$

The result is that we can eliminate the integral with the $u_x(0, \tau)$ and write the solution — the final solution— as

$$u(x, t) = \int_0^t f(\tau)\frac{x\, e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\, d\tau.$$

This form of the solution was the one given by Stokes. He wrote to Thomson:

> In working out myself various forms of the solution of the equation $dv/dt = d^2v/dx^2$ [Note: Heputs a 1 on the right hand side instead of a 1/2] under the condition $v = 0$ when $t = 0$ from$x = 0$ to $x = \infty$; $v = f(t)$ when $x = 0$ from $t = 0$ to $t = \infty$we found the solution . . .was . . .

$$v(x,t) = \frac{x}{2\sqrt{\pi}} \int_0^t (t - t')^{-\frac{3}{2}} e^{-\frac{x^2}{4(t-t')}} f(t')\,dt'.$$

In the following we illustrate didn't We Already Solve the Heat Equation? Our first application of Fourier series (the firstapplication of Fourier series) was to solve the heat equation. Let's recall the setup and the form of thesolution. We heat a circle, which we consider to be the interval $0 \le x \le 1$ with the endpoints identified.

If the initial distribution of temperature is the function $f(x)$ then the temperature $u(x,t)$ at a point $x$attime $t > 0$ is given by

$$u(x,t) = \int_0^1 g(x - y)f(y)dy\,,$$

where

$$g(u) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n u}.$$

That was our first encounter with convolution. Now, analogous to what we did, above, we might writeinstead

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}$$

and the solution as

$$u(x,t) = g(x,t) * f(x) = \int_0^1 \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n(x-y)} f(y)dy,$$

a convolution in the spatial variable, but with limits of integration just from 0 to 1. Here $f(x), g(x,t)$, and$u(x,t)$ are periodic of period 1 in x.How does this compare to what we did for the rod? If we imagine initially heating up a circle as heatingup an infinite rod by a periodic function $f(x)$ then shouldn't we be able to express the temperature $u(x,t)$for the circle as we did for the rod? We will show that the solution for a circle does have the same formas the solution for the infinite rod by means of the remarkable identity:

$$\sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} = \sqrt{2\pi t} \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}$$

Needless to say, this is not obvious.

As an aside, for general interest, a special case of this identity is particularly famous. TheJacobi theta function is defined by

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t},$$

for $t > 0$. It comes up in surprisingly diverse pure and applied fields, including number theory, and statistical mechanics (where it is used to study "partition functions"). Jacobi's identity is

$$\vartheta(t) = \frac{1}{\sqrt{t}} \vartheta\left(\frac{1}{t}\right).$$

It follows from the identity above, with $x = 0$ and replacing t by $1/2\pi t$. Applying the identity to Green's function $g(x,t)$ for heat flow on the circle we have

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x} = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t}$$

Regard the initial distribution of heat $f(x)$ as being defined on all of $\mathbf{R}$ and having period 1. Then

$$u(x,t) = \int_0^1 \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n(x-y)} f(y) dy$$

$$= \frac{1}{\sqrt{2\pi t}} \int_0^1 \sum_{n=-\infty}^{\infty} e^{-(x-y-n)^2/2t} f(y) dy \text{ (using the Green's function identity)}$$

$$= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_0^1 e^{-(x-y-n)^2/2t} f(y) dy$$

$$= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_n^{n+1} e^{-(x-u)^2/2t} f(u-n) du \text{ (substituting } u = y + n)$$

$$= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_n^{n+1} e^{-(x-u)^2/2t} f(u) du \text{ (using that } f \text{ has period 1)}$$

$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-u)^2/2t} f(u) du.$$

Incidentally, since the problem was originally formulated for heating a circle, the function $u(x,t)$ is periodic in x. Can we see that from this form of the solution? Yes, for

$$u(x+1,t) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x+1-u)^2/2t} f(u) du$$

$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-w)^2/2t} f(w+1) dw \text{ (substituting} w = u - 1)$$

$$= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-w)^2/2t} f(w) dw \left(\text{using the periodicity of} f(x) \right)$$

$$= u(x,t).$$

Now let's derive the identity

$$\sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} = \sqrt{2\pi t} \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}$$

This is a great combination of many of the things we've developed to this point, and it will come up again.

Consider the left hand side as a function of x, say

$$h(x) = \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t}.$$

This is a periodic function of period 1 — it's the periodization of the Gaussian $e^{-x^2/2t}$. (It's even not hardto show that the series converges, etc..) What are its Fourier coefficients?

We can calculate them:

$$\hat{h}(k) = \int_0^1 h(x) e^{-2\pi i k x} dx = \int_0^1 \left(\sum_{n=-\infty}^{\infty} e^{-\frac{(x-n)^2}{2t}} \right) e^{-2\pi i k x} dx$$

$$= \sum_{n=-\infty}^{\infty} \int_0^1 e^{-\frac{(x-n)^2}{2t}} e^{-2\pi i k x} dx = \sum_{n=-\infty}^{\infty} \int_{-n}^{-n+1} e^{-\frac{u^2}{2t}} e^{-2\pi i k u} du$$

$$\left(\text{substituting} u = x - n \text{ and using periodicity of} e^{-2\pi i k x} \right)$$

$$= \int_{-\infty}^{\infty} e^{-\frac{u^2}{2t}} e^{-2\pi i k u} du$$

But this last integral is exactly the Fourier transform of the Gaussian $e^{-\frac{x^2}{2t}}$ at $s = k$. We know how todo that — the answer is $\sqrt{2\pi t} e^{-2\pi^2 k^2 t}$.

We have shown that the Fourier coefficients of $h(x)$ are

$$\hat{h}(k) = \sqrt{2\pi t} e^{-2\pi^2 k^2 t}.$$

Since the function is equal to its Fourier series (really equal here because all the series converge and allthat) we conclude that

$$h(x) = \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} = \sum_{n=-\infty}^{\infty} \hat{h}(n)e^{2\pi inx} = \sqrt{2\pi t}\sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t}e^{2\pi inx} \ ,$$

and there's the identity we wanted to prove.

Now we study convolution in Action III: The Central Limit Theorem

Several times we've met the idea that convolution is a smoothing operation. Let me begin with somegraphical examples of this, convolving a discontinuous. Here areplots of this, up to $\Pi * \Pi * \Pi * \Pi$.
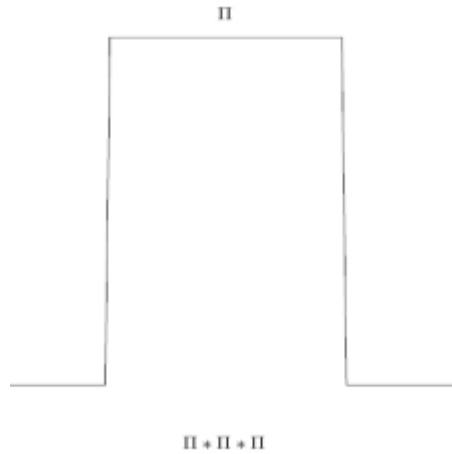
Π

Π * Π * Π

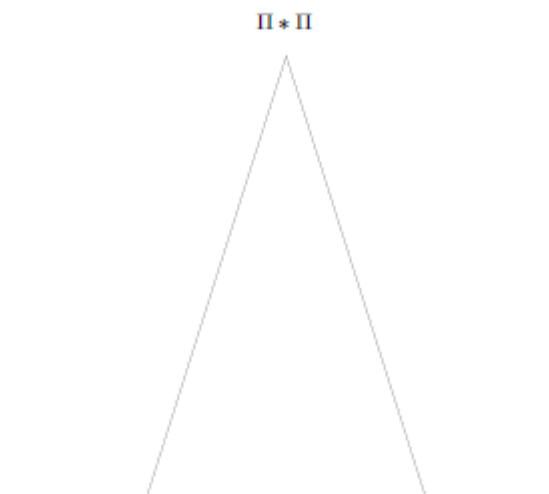**Figure (2.5)**

Π * Π

Figure (2.6)

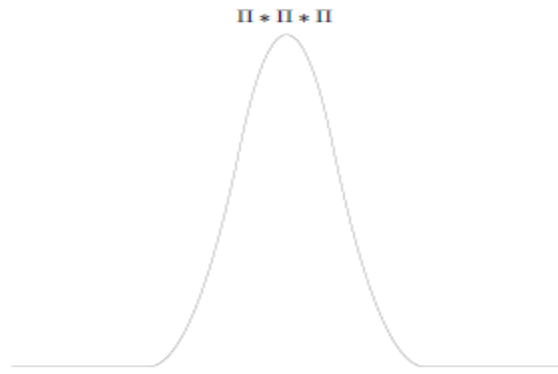$$\Pi * \Pi * \Pi$$

Figure (2.7)

$$\Pi * \Pi * \Pi * \Pi$$
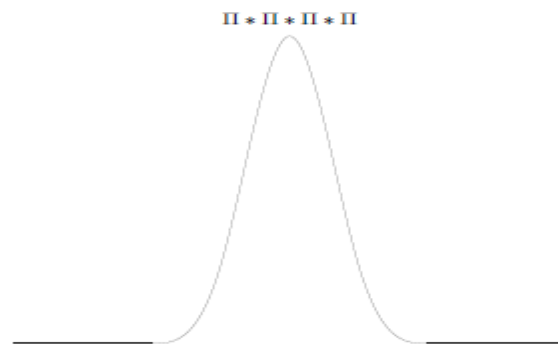

Figure (2.8)

Not only are the convolutions becoming smoother, but the unmistakable shape of a Gaussian is emerging.Is this a coincidence, based on the particularly simple nature of the function Π, or is something moregoing on? Here is a plot of, literally, a random function $f(x)$ — the values $f(x)$ are just randomly chosennumbers between 0 and 1.
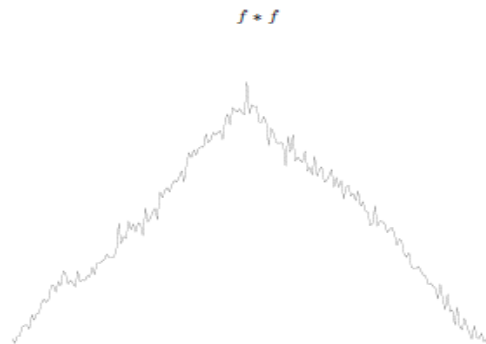
$$f * f$$

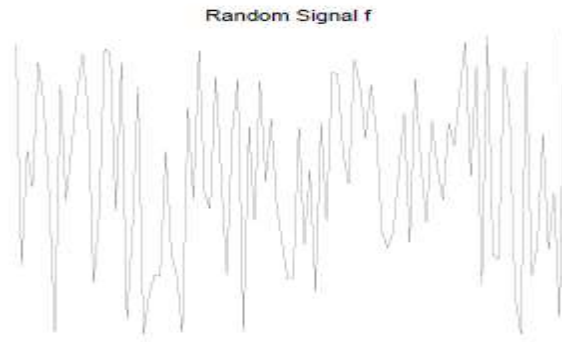**Figure (2.9)**

Random Signal f

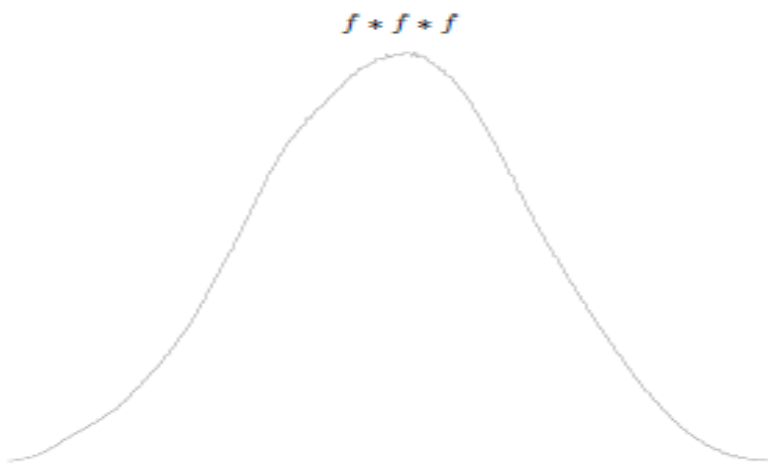**Figure (2.10)**

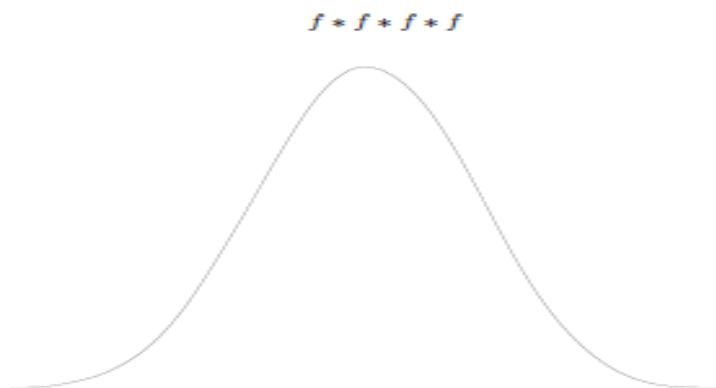$f * f * f$

Figure (2.11)

$f * f * f * f$

Figure (2.12)

From seeming chaos, again we see a Gaussian emerging. The object of this section is to explain thisphenomenon, to give substance to the following famous quotation:

> Everyone believes in the normal approximation, the experimenters because they think it is amathematical theorem, the mathematicians because they think it is an experimental fact.

G. Lippman, French Physicist, 1845–1921

The "normal approximation" (or normal distribution) is the Gaussian. The "mathematical theorem" hereis the Central Limit Theorem. To understand the theorem and to appreciate the "experimental fact", we
have to develop some ideas from probability.

In the following we discuss Random variables. In whatever field of science or engineering you pursue you will use probabilistic ideas. Youwill use theGaussian. I'm going under the assumption that you probably know some probability, and probably somestatistics, too, even if only in an informal way. For our present work, where complete generality based
on exquisitely precise terminology is not the goal, we only need a light dose of some of the fundamentalnotions.

The fundamental notion is the random variable. A random variable is a number you don't know yet. By that we mean that it, or rather its value, is the numerical result of some process, like a measurementor the result of an experiment. The assumption is that you can make the measurement, you can performthe experiment, but until you do you don't know the value of the random variable. It's called "random"because a particular object to be measured is thought of as being drawn "at random" from a collection ofall such objects. For example:

Random Variable	Value of random variable
Height of people in US population	Height of particular person
Length of pins produced	Length of particular pin
Momentum of atoms in a gas	Momentum of particular atom
Resistance of resistors off a production line	Resistance of a particular resistor
Toss of coin	0 or 1 (head or tail)
Roll of dice	Sum of numbers that come up

A common notation is to write X for the name of the random variable and x for its value. If you thenthink that a random variable X is just a function, you're right, but deciding what the domain of sucha function should be, and what mathematical structure to require of both the domain and the function,demands the kind of precision that we don't want to get into. This was a long time in

coming. Consider,for example, Mark Kac's comment: "independent random variables were to me (and others, includingmy teacher Steinhaus) shadowy and not really well-defined objects." Kac was one of the most eminentprobabilists of the 20th century.

Now we illustrate Probability distributions and probability density functions. "Random variable" is the fundamental notion, but not the fundamental object of study. For a givenrandom variable what we're most interested in is how its values are distributed. For this it's helpful todistinguish between two types of random variables.

1. A random variable is discrete if its values are among only a finite number of possibilities.

   For example "Roll the die" is a discrete random variable with values 1, 2, 3, 4, 5 or 6. "Tossthe coin" is a discrete random variable with values 0 and 1. (A random variable with values 0and 1 is the basic random variable in coding and information theory.)

2. A random variable is continuous if its values do not form a discrete set, typically filling up one ormore intervals of real numbers.

   For example "length of a pin" is a continuous random variable since, in theory, the length of apin can vary continuously.

For a discrete random variable we are used to the idea of displaying the distribution of values as a histogram.We set up bins, one corresponding to each of the possible values, we run the random process however manytimes we please, and for each bin we draw a bar with height indicating the percentage that value occursamong all actual outcomes of the runs. Since we plot percentages, or fractions, the total area of thehistogram is 100%, or just 1.

A series of runs of the same experiment or the same measurement will produce histograms of varying shapes. We often expect some kind of limiting shape as we increase the number of runs, or we maysuppose that the ideal distribution has some shape, and then compare the actual data from a series of runsto the ideal, theoretical answer.

The theoretical histogram is called the probability distribution. Andthe function that describes the histogram (the shape of the distribution) is called the probabilitydensity function or pdf, of the random variable.

Is there a difference between the probability distribution and the probability density function? No, notreally — it's like distinguishing between the graph of a function and the function. Both terms are incommon use, more or less interchangeably.

The probability that any particular value comes up is the area of its bin in the probability distribution,which is therefore a number between 0 and 1.

If the random variable is called $X$and the value we're interested in is $x$we write this as

$$\text{Prob}(X = x) = \text{ area of the bin over}x .$$

Also

$$\text{Prob}(a \le X \le b) = \text{ areas of the bins from}a \text{ to}b .$$

Thus probability is the percentage of the occurrence of a particular outcome, or range of outcomes, amongall possible outcomes. We must base the definition of probability on what we presume or assume is thedistribution function for a given random variable. A statement about probabilities for a run of experimentsis then a statement about long term trends, thought of as an approximation to the ideal distribution.

One can also introduce probability distributions and probability density functions for continuous randomvariables. You can think of this — in fact you probably should think of this — as a continuous version ofa probability histogram. It's a tricky business, however, to "take a limit" of the distribution for a discreterandom variable, which have bins of a definite size, to produce a distribution for a continuous randomvariable, imagining the latter as having infinitely many infinitesimal bins.

It's easiest, and best, to define the distribution for a continuous random variable directly.

A probability density function is a nonnegative function $p(x)$ with area 1, i.e.,

$$\int_{-\infty}^{\infty} p(x)dx = 1.$$

Remember, $x$is the measured value of some experiment. By convention, we take $x$to go from $-\infty$to ∞ so we don't constantly have to say how far the values extend.

Here's one quick and important property of pdfs:

1. If $p(x)$ is a pdf and $a > 0$ then $ap(ax)$ is also a pdf.

    To show this we have to check that the integral of $ap(ax)$ is 1. But

    $$\int_{-\infty}^{\infty} a\, p(ax)dx = \int_{-\infty}^{\infty} a\, p(u)\frac{1}{a}\, du = \int_{-\infty}^{\infty} p(u)du = 1,$$

    making the change of variable $u = ax$. We'll soon see this property in action.

2. We think of a pdf as being associated with a random variable Xwhose values are $x$and we write $p_X$if we want to emphasize this. The

(probability) distribution of X is the graph of p_X, but, again, the terms probability density function and probability distribution are used interchangeably.

3. Probability is defined by

$$\text{Prob}(X \le a) = \text{ Area under the curve for} x \le a$$

$$= \int_{-\infty}^{a} p_X(x)dx \, ,$$

Also

$$\text{Prob}(a \le X \le b) = \int_{a}^{b} p_X(x)dx \, .$$

For continuous random variables it really only makes sense to talk about the probability of a range of values occurring, not the probability of the occurrence of a single value. Think of the pdf as describing a limit of a (discrete) histogram: If the bins are becoming infinitely thin, what kind of event could land in an infinitely thin bin?

Finally, for variable t, say, we can view

$$P(t) = \int_{-\infty}^{t} p(x)dx$$

as the "probability function". It's also called the cumulative probability or the cumulative density function. We then have

$$\text{Prob}(X \le t) = P(t)$$

and

$$\text{Prob}(a \le X \le b) = P(b) - P(a).$$

According to the fundamental theorem of calculus we can recover the probability density function from $P(t)$ by differentiation:

$$\frac{d}{dt}P(t) = p(t).$$

In short, to know $p(t)$ is to know $P(t)$ and vice versa. You might not think this news is of any particular practical importance, but you're about to see that it is.

Now we discuss mean, variance, and standard deviation. Suppose X is a random variable with pdf $p(x)$. The x's are the values assumed by X, so the mean μ of X is the weighted average of these values, weighted according to p. That is,

$$\mu(X) = \int_{-\infty}^{\infty} xp(x)dx \, .$$

Be careful here — the mean of X, defined to be the integral of $xp(x)$, is not the average value of the function $p(x)$. It might be that $\mu(X) = \infty$, of course, i.e., that the integral of $xp_X(x)$ does not converge.

If $\mu(X) < \infty$ then we can always "subtract off the mean" to assume that X has mean zero. Here's what this means, no pun intended; in fact let's do something slightly more general. What do we mean by $X - a$, when X is a random variable and a is a constant? Nothing deep — you "do the experiment" to get a value of X (X is a number you don't know yet) then you subtract a from it. What is the pdf of $X - a$? To figure that out, we have

$$\text{Prob}(X - a \le t) = \text{Prob}(X \le t + a) = \int_{-\infty}^{t+a} p(x)dx$$

$$= \int_{-\infty}^{t} p(u + a)du \text{ (substituting } u = x - a).$$

This identifies the pdf of $X - a$ as $p(x + a)$, the shifted pdf of X.

Next, what is the mean of $X - a$. It must be $\mu(X) - a$ (common sense, please). Let's check this now knowing what pdf to integrate.

$$\mu(X - a) = \int_{-\infty}^{\infty} xp(x + a)dx = \int_{-\infty}^{\infty} (u - a)p(u)du \text{ (substituting } u = x + a)$$

$$= \int_{-\infty}^{\infty} up(u)du - a \int_{-\infty}^{\infty} p(u)du = \mu(X) - a.$$

Note that translating the pdf $p(x)$ to $p(x + a)$ does nothing to the shape, or areas, of the distribution, hence does nothing to calculating any probabilities based on $p(x)$. As promised, the mean is $\mu(X) - a$. We are also happy to be certain now that "subtracting off the mean", as in $X - \mu(X)$, really does result in a random variable with mean 0. This normalization is often a convenient one to make in deriving formulas.

Suppose that the mean $\mu(X)$ is finite. The variance σ^2 is a measure of the amount that the values of the random variable deviate from the mean, on average, i.e., as weighted by the pdf $p(x)$. Since some values are above the mean and some are below we weight the square of the differences, $(x - \mu(X))^2$, by $p(x)$ and define

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x - \mu(X))^2 p(x)dx.$$

If we have normalized so that the mean is zero this becomes simply

$$\sigma^2(X) = \int_{-\infty}^{\infty} x^2 p(x)dx.$$

The standard deviation is $\sigma(X)$, the square root of the variance. Even if the mean is finite it might be that $\sigma^2(X)$ is infinite; this, too, has to be checked for any particular example.

We've just seen that we can normalize the mean of a random variable to be 0. Assuming that the varianceis finite, can we normalize it in some helpful way? Suppose Xhas pdf $p$and let a be a positive constant.
Then

$$\text{Prob}\left(\frac{1}{a}X \leq t\right) = \text{Prob}(X \leq at) = \int_{-\infty}^{at} p(x)dx$$

$$= \int_{-\infty}^{t} ap(au)du \left(\text{making the substitution} u = \frac{1}{a}x\right)$$

This says that the random variable $\frac{1}{a}X$has pdf $ap(ax)$. (Here in action is the scaled pdf $ap(ax)$, which wehad as an example of operations on pdf's.)
Suppose that we've normalized the mean of $X$to be 0. Thenthe variance of $\frac{1}{a}X$is

$$\sigma^2\left(\frac{1}{a}X\right) = \int_{-\infty}^{\infty} x^2 ap(ax)dx$$

$$= a\int_{-\infty}^{\infty}\frac{1}{a^2}u^2p(u)\frac{1}{a}du \text{ (making the substitution} u = ax)$$

$$= \frac{1}{a^2}\int_{-\infty}^{\infty} u^2p(u)du = \frac{1}{a^2}\sigma^2(X)$$

In particular, if we choose $a = \sigma(X)$ then the variance of $\frac{1}{a}X$is one. This is also a convenient normalizationfor many formulas.

In summary:

Given a random variable $X$with finite $\mu(X)$ and $\sigma(X) < \infty$, it is possible to normalize and assumethat $\mu(X) = 0$ and $\sigma^2(X) = 1$.

In the following we discuss two examples:Let's be sure we have two leading examples of pdfs to refer to.

1. The uniform distribution "Uniform" refers to a random process where all possible outcomes are equally likely. In the discrete case tossing a coin or throwing a die are examples. All bins in the ideal histogram have the same height, two bins of height $1/2$ for the toss of a coin, six bins of height $1/6$ for the throw of a single die, and Nbins of height $1/N$for a discrete random variable with Nvalues.

For a continuous random variable the uniform distribution is identically 1 on an interval of length 1 andzero elsewhere. If we shift to the interval from $-1/2$ to $1/2$, it's thegraph of the ever versatile rect function. $\Pi(x)$ is now starring in yet another role, that of the uniformdistribution.

The mean is 0, obviously, but to verify this formally:

$$\mu = \int_{-\infty}^{\infty} x\Pi(x)dx = \int_{-1/2}^{1/2} xdx = \frac{1}{2}x^2\Big]_{-1/2}^{+1/2} = 0.$$

The variance is then

$$\sigma^2 = \int_{-\infty}^{\infty} x^2\Pi(x)dx = \int_{-1/2}^{1/2} x^2dx = \frac{1}{3}x^3\Big]_{-1/2}^{+1/2} = \frac{1}{12},$$

perhaps not quite so obvious.

2. The normal distribution:

So it seems appropriatethat at some point I mention:

The Gaussian is a pdf.

Indeed, the Gaussian

$$g(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}.$$

is a pdf with mean $\mu$and variance σ^2. The distribution associated with such a Gaussian is called a normaldistribution.

Now we study Independence. An important extra property that random variables may have is independence. The plain English description of independence is that one event or measurement doesn't influence another event or measurement.

Each flip of a coin, roll of a die, or measurement of a resistor is a new event, not influenced by previousevents.

Operationally, independence implies that the probabilities multiply: If two random variables X_1 and $X_2$are independent then

$$\text{Prob}(X_1 \le a \text{ and} X_2 \le b) = \text{Prob}(X_1 \le a) \cdot \text{Prob}(X_2 \le b).$$

In words, if $X_1 \le a$occurs r percent and $X_2 \le b$occurs s percent then, if the events are independent, thepercent that $X_1 \le a$occurs and $X_2 \le b$occurs is $r$percent of $s$percent, or rspercent.

In the following we illustrate Convolution appears. Using the terminology we've developed, we can begin to be more precise about the content of the CentralLimit Theorem. That result— the ubiquity of the bell-shaped curve — has to do with sums of independentrandom variables and with the distributions of those sums.

While we'll work with continuous random variables, let's look at the discrete random variable X = "rollthe dice" as an example. The ideal histogram for the toss of a single die is uniform — each number 1through 6 comes up with equal probability. We might represent it pictorially like this:
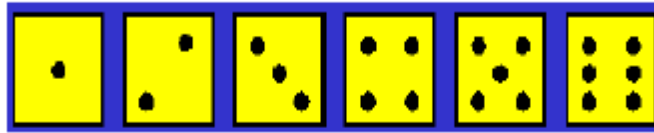
Figure (2.13)

We don't mean to think just of a picture of dice here — I mean to think of the distribution as six bins ofequal height $1/6$, each bin corresponding to one of the six possible tosses.

What about the sum of the tosses of two dice? What is the distribution, theoretically, of the sums? Thepossible values of the sum are 2 through 12, but the values do not occur with equal probability. There'sonly one way of making 2 and one way of making 12, but there are more ways of making the other possiblesums. In fact, 7 is the most probable sum, with six ways it can be achieved. We might represent thedistribution for the sum of two dice pictorially like this:
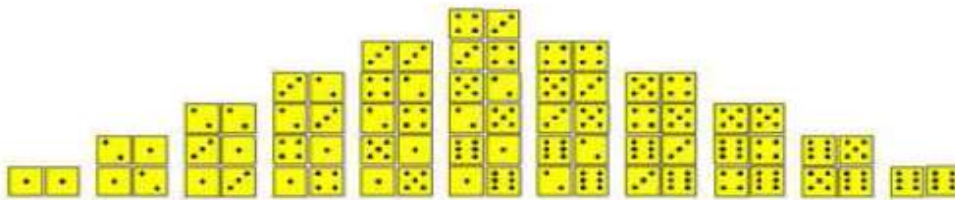

Figure (2.14)

It's triangular. Now let's see , For the single random variable $X =$ "roll one die" we have a distributionlike a rect function. For the sum, say random variables $X_1 + X_2 =$ "roll of die 1 plus roll of die 2", thedistribution looks like the triangle function . .

The key discovery is this:

Convolution and probability density functions. The probability density function of thesum of two independent random variables is the convolution of the probability density functionsof each.

We can get a good intuitive sense of why this result might hold by looking again at the discrete case andat the example of tossing two dice. To ask about the distribution of the sum of two dice is to ask aboutthe probabilities of particular numbers coming up, and these we can compute directly using the rules ofprobability. Take, for example, the probability that the sum is 7. Count the ways, distinguishing whichthrow is first:

Prob($Sum = 7$)

= Prob({1 and 6}or{2 and 5}or{3 and 4}or{4 and 3}or{5 and 2}or{6 and 1})

= Prob(1 and 6) + Prob(2 and 5) + Prob(3 and 4) + Prob(4 and 3)

+ Prob(5 and 2)

+ Prob(6 and 1)(probabilities add when events are mutually exclusive)

= Prob(1)Prob(6) + Prob(2)Prob(5) + Prob(3)Prob(4) + Prob(4)Prob(3)

+ Prob(5)Prob(2)

+ Prob(6)Prob(1)(probabilities multiply when events are independent)

$$= 6\left(\frac{1}{6}\right)^2 = \frac{1}{6}.$$

The particular answer, Prob($Sum = 7$) = 1/6, is not important here — it's the form of the expressionfor the solution that should catch your eye. We can write it as

$$\text{Prob}(Sum = 7) = \sum_{k=1}^{6} \text{Prob}(k)\text{Prob}(7 - k)$$

which is visibly a discrete convolution of Prob with itself — it has the same form as an integral convolutionwith the sum replacing the integral.

We can extend this observation by introducing

$$p(n) = \begin{cases} \dfrac{1}{6} & n = 1, 2, \dots, 6 \\ 0 & \text{otherwise} \end{cases}$$

This is the discrete uniform density for the random variable "Throw one die". Then, by the same reasoningas above,

$$\text{Prob}(\text{sum of two dice} = n) = \sum_{k=-\infty}^{\infty} p(k)p(n - k).$$

one can check that this gives the right answers, including the answer 0 for n bigger than 12 or n lessthan 2:

n	Prob (sum = n)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/35
10	3/36
11	2/36
12	1/36

Now let's turn to the case of continuous random variables, and in the following argument look for similaritiesto the example we just treated. Let X_1 and X_2 be independent variables with probability densityfunctions $p_1(x_1)$ and $p_2(x_2)$. Because X_1 and X_2 are independent,

$$\text{Prob}(a_1 \leq X_1 \leq b_1 \text{ and } a_2 \leq X_2 \leq b_2) = \left(\int_{a_1}^{b_1} p_1(x_1) d\,x_1 \right) \left(\int_{a_2}^{b_2} p_2(x_2) d\,x_2 \right)$$

Using what has now become a familiar trick, we write this as a double integral.

$$\left(\int_{a_1}^{b_1} p_1(x_1) d\,x_1 \right) \left(\int_{a_2}^{b_2} p_2(x_2) d\,x_2 \right) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} p_1(x_1) p_2(x_2) d\,x_1 d\,x_2,$$

that is,

$$\text{Prob}(a_1 \leq X_1 \leq b_1 \text{ and } a_2 \leq X_2 \leq b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} p_1(x_1) p_2(x_2) d\,x_1 d\,x_2.$$

If we let a_1 and a_2 drop to $-\infty$ then

$$\text{Prob}(X_1 \leq b_1 \text{ and } X_2 \leq b_2) = \int_{-\infty}^{b_2} \int_{-\infty}^{b_1} p_1(x_1) p_2(x_2) d\,x_1 d\,x_2.$$

Since this holds for any b1 and b2, we can conclude that

$$\text{Prob}(X_1 + X_2 \leq t) = \iint_{x_1+x_2 \leq t} p_1(x_1) p_2(x_2) d\,x_1 d\,x_2$$

for every t. In words, the probability that $X_1 + X_2 \leq t$ is computed by integrating the joint probabilitydensity $p_1(x_1)p_2(x_2)$ over the region in the (x_1, x_2)-plane where $x_1 + x_2 \leq t$.

We're going to make a change of variable in this double integral. We let

$$x_1 = u$$
$$x_2 = v - u$$

Notice that $x_1 + x_2 = v$. Thus under this transformation the (oblique) line $x_1 + x_2 = t$ becomes thehorizontal line $v = t$, and the region $x_1 + x_2 \leq t$ in the (x_1, x_2)-plane becomes the half-plane $v \leq t$ in the (u, v)-plane.
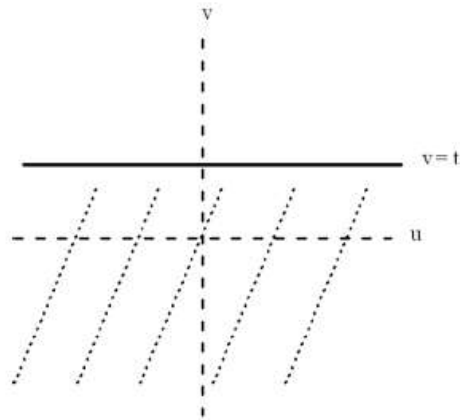
Figure (2.15)
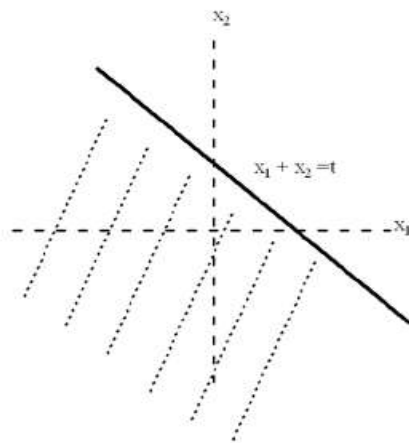
Figure (2.16)

The integral then becomes

$$\iint_{x_1+x_2\leq t} p_1(x_1)p_2(x_2)d\,x_1 d\,x_2 = \int_{-\infty}^{t}\int_{-\infty}^{\infty} p_1(u)p_2\,(v-u)d\,u\,d\,v$$

(the convolution of p_1 and p_2 is inside!)

$$= \int_{-\infty}^{t} (p_2 * p_1)(v)dv.$$

To summarize, we now see that the probability $\text{Prob}(X_1 + X_2 \leq t)$ for any t is given by

$$\text{Prob}(X_1 + X_2 \leq t) = \int_{-\infty}^{t} (p_2 * p_1)(v)dv.$$

Therefore the probability density function of $X_1 + X_2$ is $(p_2 * p_1)(t)$.

This extends to the sum of any finite number of random variables: If $X_1, X_2, ..., X_n$ are independentrandom variables with probability density

76

functions $p_1, p_2, \ldots, p_n$, respectively, then the probability densityfunction of $X_1 + X_2 + \cdots + X_n$ is $p_1 * p_2 * \ldots * p_n$. Cool. Cool. . . . Cool.

For a single probability density$p(x)$ we'll write

$$p^{*n}(x) = (p * p * \ldots * p)(x)$$

$(n$ factors of$p$, i.e., $n-1$convolutions of$p$ with itself).

In the following we discuss The Central Limit Theorem: The Bell Curve Tolls for Thee. The Central Limit Theorem says something like the sum of $n$independent random variables is well approximated by a Gaussian if $n$is large. That means the sum is distributed like a Gaussian. To make a truestatement, we have to make a few assumptions — but not many — on how the random variables themselves are distributed. Call the random variables $X_1, X_2, \ldots, X_n$. We assume first of all that the X's areindependent. We also assume that all of X's have the same probability density function. There's someterminology and an acronym that goes along with this, naturally. One says that the X's are independentand identically distributed, or iid. In particular the X's all have the same mean, say μ, and they all havethe same standard deviation, say σ.

Consider the sum

$$S_n = X_1 + X_2 + \cdots + X_n.$$

We want to say that S_n is distributed like a Gaussian as $n$increases, but which Gaussian? The mean andstandard deviation for the X's are all the same, but for S_n they are changing with n. It's not hard toshow, though, that for S_n the mean scales by n and thus the standard deviation scales by$\sqrt{n}$:

$$\mu(S_n) = n\mu$$
$$\sigma(S_n) = \sqrt{n}\,\sigma$$

So to make sense of S_n approaching a particular Gaussian we should therefore recenter and rescale thesum, say fix the mean to be zero, and fix the standard deviation to be 1. That is, we should work with

$$\frac{S_n - n\mu}{\sqrt{n}\sigma}$$

and ask what happens as $n \to \infty$. One form of the Central Limit Theorem says that

$$\lim_{n \to \infty} \text{Prob}\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

On the right hand side is the Gaussian $(1/\sqrt{2\pi})e^{-x^2/2}$ with mean 0 and standard deviation 1. Thetheorem says that probabilities for the normalized sum of the random variables approach those based onthis Gaussian.

We'll focus on the convergence of the pdf's for S_n — sort of an unintegrated form of the way the CLT isstated, above. Let $p(x)$ be the common probability density function for the $X_1, X_2, \ldots, X_n$. (The pdf for the iidX's, for those who like to compress their terminology.) We'll start by assuming already that $\mu = 0$ and $\sigma = 1$ for the X's. This means that

$$\int_{-\infty}^{\infty} xp(x)dx = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2p(x)dx = 1,$$

in addition to

$$\int_{-\infty}^{\infty} p(x)dx = 1,$$

which is true for every pdf.

Now, the mean of S_n is zero, but the standard deviation is $\sqrt{n}$, so we want to work $S_n/\sqrt{n}$. What is thepdf of this? We've shown that the pdf for $S_n = X_1 + \cdots + X_n$ is

$$p^{*n}(x) = (p * p * \ldots * p)(x).$$

Hence the probability density function for $S_n/\sqrt{n}$is

$$p_n(x) = \sqrt{n}p^{*n}(\sqrt{n}\,x).$$

(Careful here: It's $(p * p * \ldots * p)(\sqrt{n}\,x)$, not $p(\sqrt{n}\,x) * p(\sqrt{n}\,x) * \ldots p(\sqrt{n}\,x)$.)

We're all set to show:

Theorem (2.2.5): (Central Limit Theorem)

Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed randomvariables with mean 0 and standard deviation 1. Let $p_n(x)$ be the probability density functionfor $\frac{S_n}{\sqrt{n}} = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}$. Then

$$p_n(x) \to \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\text{as } n \to \infty.$$

The idea is to take the Fourier transform of p_n, which, by the Convolution Theorem, will essentially bethe product of the Fourier transforms of p. Products are easier than convolutions, and the hope is to usethe assumptions on $p$to get some information on the form of this product as $n \to \infty$.

Begin with the Fourier transform of

$$p_n(x) = \sqrt{n}p^{*n}(\sqrt{n}\,x).$$

We'll use the capital letter notation and write $P(s) = \mathcal{F}p(s)$. Then the Fourier transform of $p_n(x)$ is

$$p^n(s/\sqrt{n})\text{(ordinary nth power here)}.$$

The normalization of mean zero and standard deviation 1 allows us to dosomething with $P(s/\sqrt{n})$. Usinga Taylor series approximation for the exponential function, we have

$$P(s/\sqrt{n}) = \int_{-\infty}^{\infty} e^{-\frac{2\pi isx}{\sqrt{n}}} p(x)dx$$

$$= \int_{-\infty}^{\infty} \left(1 - \frac{2\pi isx}{\sqrt{n}} + \frac{1}{2}\left(\frac{2\pi isx}{\sqrt{n}}\right)^2 + \text{small}\right) p(x)dx$$

$$= \int_{-\infty}^{\infty} \left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2 s^2 x^2}{n} + \text{small}\right) p(x)dx$$

$$= \int_{-\infty}^{\infty} p(x)dx - \frac{2\pi is}{\sqrt{n}} \int_{-\infty}^{\infty} xp(x)dx - \frac{2\pi^2 s^2}{n} \int_{-\infty}^{\infty} x^2 p(x)dx$$

$$+ \int_{-\infty}^{\infty} (\text{small})\, p(x)dx = 1 - \frac{2\pi^2 s^2}{n} + \text{small}.$$

In the last step we used the normalizations

$$\int_{-\infty}^{\infty} p(x)dx = 1, \quad \int_{-\infty}^{\infty} x\, p(x)dx = 0, \quad \int_{-\infty}^{\infty} x^2\, p(x)dx = 1.$$

That "small" term tends to 0 faster than $1/n$as $n \to \infty$.
Using the well known fact that$(1 + x/n)^2 \to e^x$, we have for large n

$$P^n\left(\frac{s}{\sqrt{n}}\right) \approx \left(1 - \frac{2\pi^2 s^2}{n}\right)^n \approx e^{-2\pi^2 s^2}.$$

Taking the inverse Fourier transform of $e^{2\pi^2 s^2}$and knowing what happens to the Gaussian, taking thelimit as $n \to \infty$, taking the rest of the day off for a job well done, we conclude that

$$p_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Now we are discuss Fourier transform formulas under different normalizations. With convolution now part of our working lives we've seen the major operations and formulas involvingFourier transforms. As above we cautioned that there are different conventions for definingthe Fourier transform, and different conventions result in different formulas. Here is a summary of whatyou'll find out there.
To be as general as possible let's write, as we did back,

$$\mathcal{F}f(s) = \frac{1}{A} \int_{-\infty}^{\infty} e^{iBst} f(t)dt.$$

We use $A = 1$ and $B = -2\pi$but different Fourier practitioners may well use any of the following pairs ofvalues:

$$A = \sqrt{2\pi} \qquad B = \pm1$$
$$A = 1 \qquad B = \pm2\pi$$
$$A = 1 \qquad B = \pm1$$

Whatever you choose, here's what you'll then have to live with:

$$\mathcal{F}\mathcal{F}f = \frac{2\pi}{A^2|B|}f^{-} \qquad \mathcal{F}(f') = -iB\mathcal{F}f$$
$$(\mathcal{F}f)' = iB\mathcal{F}(tf(t))$$
$$\mathcal{F}(f * g) = A(\mathcal{F}f)(\mathcal{F}g).$$

Chapter 3

A physical Analogy For Distributions

Section (1. 3): Right Functions for Fourier Transforms and little on integrals

We begin this section by studying the Day of Reckoning .we've been playing a little fast and loose with the Fourier transform − applying Fourier inversion, appealing to duality, and all that . "fast and loose" is an understatement if ever there was one, but it's also true that we haven't done anything "wrong" . all of our formulas and all of our applications have been correct, if not fully justified. Nevertheless, we have to come to terms with some fundamental questions. It well take us some time, but in the end we will have settled on a very wide class of signals with these properties:

1. The allowed signals include δ's, unit steps, ramps, $sines, cosines,$ and all other standard signals that the world's economy depends on.
2. The Fourier transform and its inverse are defined for all of these signals.
3. Fourier inversion work.
 These are the three most important features of the development to come, but we'll also reestablish some of our specific results and as an added benefit we'll even finish off differential calculus.
Now we study A too simple criterion and an example. It's not hard to write down an assumption on a function that guarantees the existence of its Fourier transform and even implies a little more than existence.
If $\int_{-\infty}^{\infty}|f(t)|dt < \infty$ then $\mathcal{F}f$ and $\mathcal{F}^{-1}f$ exist and are continuous.
Existence follows from

$$|\mathcal{F} f \, (s)| = |\int_{-\infty}^{\infty} e^{-2\pi i s t} f(t)dt \,|$$

$$\leq \int_{-\infty}^{\infty} |e^{-2\pi i s t}|| f(t)|dt = \int_{-\infty}^{\infty}| f(t)|dt < \infty.$$

Here we've used that the magnitude of the integral is less that the integral of the magnitude.
Continuity is the little extra information we get beyond existence. Continuity
 Follows as follows. For any s and s' we have

$$|\mathcal{F} f \, (s) - \mathcal{F} f \, (s')| = |\int_{-\infty}^{\infty} e^{-2\pi i s t} f(t)dt \, - \, \int_{-\infty}^{\infty} e^{-2\pi i s' t} f(t)dt \,|$$

$$=\left|\int_{-\infty}^{\infty}(e^{-2\pi ist}-e^{-2\pi is't})f(t)dt\right|\leq\int_{-\infty}^{\infty}|e^{-2\pi ist}-e^{-2\pi is't}||f(t)|dt$$

As a consequence of $\int_{-\infty}^{\infty}|f(t)|dt<\infty$ we can take the limit as $s'\rightarrow s$ inside the integral. if we do that then $|e^{-2\pi ist}-e^{-2\pi is't}|\rightarrow 0$, that is,

$|\mathcal{F}f(s)-\mathcal{F}f(s')|\rightarrow 0$ as $s'\rightarrow s$

Which says that $\mathcal{F}f(s)$ is continuous. The same argument works to show that $\mathcal{F}^{-1}f$ is continuous.

We haven't said anything here about Fourier inversion – no such statement appears in the criterion. Let's look right away at an example.

The very first example we computed, and still an important one, is the Fourier transform of Π. We found directly that

$$\mathcal{F}\Pi(s)=\int_{-\infty}^{\infty}e^{-2\pi ist}\Pi(t)dt=\int_{-1/2}^{1/2}e^{-2\pi ist}\,dt=sinc\,s.$$

No problem there, no problem whatsoever. The criterion even applies; Π is in $L^1(R)$

Since

$$\int_{-\infty}^{\infty}|\Pi(t)|dt=\int_{-1/2}^{1/2}1\,dt=1.$$

Furthermore, the transform $\mathcal{F}\Pi(s)=sinc\,s$ is continuous. That's worth remarking on: Although the signal jumps (Π has discontinuity) the Fourier transform does not,

Just as guaranteed by the preceding result – make this part of your intuition on the Fourier transform vis a vis the signal.

Appealing to the Fourier inversion theorem and what we called duality, we then said

$$\mathcal{F}sinc(t)=\int_{-\infty}^{\infty}e^{-2\pi ist}\,sinc\,t\,dt=\Pi(s).$$

Here we have a problem. The sinc function does not satisfy the integrability criterion. It is my sad duty to inform you that $\int_{-\infty}^{\infty}|sinc\,t|\,dt=\infty.$

We'll give you two ways of seeing the failure of $|sinc\,t|$ to be integrable. First , if sinc did satisfy the criterion $\int_{-\infty}^{\infty}|sinc\,t|dt<\infty$ then its Fourier transform

would be continuous .but its Fourier transform, which has to come out to be Π, is not continuous . or, if we don't like that, here's a direct argument. We can find infinitely

Many intervals where $|sin\,\pi t| \geq \frac{1}{2}$; this happens when t is between 1/6 and 5/6, and that repeats for infinitely many intervals, for example on $I_n = [1/6 + 2n, 5/6 + 2n], n = 0,1,2, \ldots$, because $sin\pi t$ is periodic of period 2. The I_n all have length 2/3. On I_n we have $|t| \leq 5/6 + 2n$, so

$$\frac{1}{|t|} \geq \frac{1}{\frac{5}{6} + 2n}$$

And

$$\int_{I_n} \frac{|sin\pi t|}{\pi|t|}\, dt \geq \frac{1}{2\pi}\frac{1}{\frac{5}{6} + 2n}\int_{I_n} dt = \frac{1}{2\pi}\frac{2}{3}\frac{1}{\frac{5}{6} + 2n}.$$

Then

$$\int_{-\infty}^{\infty} \frac{|sin\pi t|}{\pi|t|}\, dt \geq \sum_n \int_{I_n} \frac{|sin\pi t|}{\pi|t|} = \frac{1}{3\pi}\sum_{n=1}^{\infty} \frac{1}{\frac{5}{6} + 2n} = \infty.$$

It's true that $|sinc\, t\,| = |sin\,\pi t/\pi t|\ tends\ to\ o\ as\ t \rightarrow \pm\infty - the\ 1/t$ factor makes that happen – but not "fast enough" to make the integral of $|sinc\, t\,|$converge.

This is most basic example in the theory. It's not clear that the integral defining the Fourier transform of $sinc$ exists, at least it doesn't follow from the criterion. Doesn't this bother you? Isn't it a little embarrassing that multibillion dollar industries seem to depend on integrals that don't converge?

In fact, there isn't so much of a problem with either Π or sinc. It is true that

$$\int_{-\infty}^{\infty} e^{-2\pi ist}\, sinc\, s\, ds = \begin{cases} 1 & |t| < \dfrac{1}{2} \\ 0 & |t| > \dfrac{1}{2} \end{cases}$$

However showing this – evaluating the improper integral that defines the Fourier transform – requires special arguments and techniques. The $sinc$ function oscillates, as do the real and imaginary parts of the complex exponential, and integrating $e^{-2\pi ist}\ sinc\ s$ involves enough cancellation for the limit

$$\lim_{\substack{a\rightarrow-\infty \\ b\rightarrow\infty}} \int_a^b e^{-2\pi ist}\, sinc\, s\, ds$$

To exist.

Thus Fourier inversion, and duality, can be pushed through in this case. At least
Almost .you'll notice that we didn't say anything about the points $t = \pm 1/2$,
where there's a jump in Π in the time domain. In those cases the improper
integral does not exist, but with some additional interpretations one might be
able to convince a sympathetic friend that

$$\int_{-\infty}^{\infty} e^{-2\pi i(\pm 1/2)s} \, sinc \, s \, ds = \frac{1}{2}$$

In the appropriate sense (invoking "principle value integrals").at best this is post
hoc and needs some fast talking.

The truth is that cancellations that occur in the $sinc$ integral or in its Fourier
transform are a very subtle and dicey thing. Such risky encounters are to
avoided.

We'd like a more robust, trustworthy theory.

The news so far.Here's a quick summary of the situation. The Fourier transform
of $f(t)$ is definedwhen

$$\int_{-\infty}^{\infty} |f(t)| dt \; < \; \infty.$$

We allow f to be complex valued in this definition. The collection of all
functions on **R** satisfying thiscondition is denoted by $L^1(\textbf{R})$, the superscript 1
indicating that we integrate $|f(t)|$ to the first power.The L^1-norm of F is
defined by

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(t)| dt.$$

Many of the examples we worked with are L^1-functions — the rect function, the
triangle function, theexponential decay (one or two-sided), Gaussians — so our
computations of the Fourier transforms in thosecases were perfectly justifiable
(and correct). Note that L^1-functions can have discontinuities, as in therect
function.

The criterion says that *if $f \in L^1(\textbf{R})$ then $\mathcal{F}f$* exists. We can also say

$$|\mathcal{F}f(s)| = |\int_{-\infty}^{\infty} e^{-2\pi i st} f(t) dt| \leq \int_{-\infty}^{\infty} |f(t)| dt = \|f\|_1.$$

That is:

I. The magnitude of the Fourier transform is bounded by the L^1-norm of the function.

This is a handy estimate to be able to write down — we'll use it shortly. However, to issue a warning:

Fourier transforms of $L^1(\mathbf{R})$ functions may themselves notbe in L^1, like for the $sinc$ function, so we don't know without further work what more can be done, if anything.

The conclusion is that L^1-integrability of a signal is just too simple a criterion on which to build a reallyhelpful theory. This is a serious issue for us to understand. Its resolution will greatly extend the usefulnessof the methods we have come to rely on.

There are other problems, too. Take, for example, the signal $f(t) = cos2\pi t$. As it stands now, this signaldoes not even havea Fourier transform — does not have a spectrum! — for the integral

$$\int_{-\infty}^{\infty} e^{-2\pi ist} \, cos2\pi t \, dt$$

does not converge, no way, no how. This is no good.

Before we bury $L^1(\mathbf{R})$ as too restrictive for our needs, here's one more good thing about it. There's actuallya stronger consequence for $\mathcal{F}f$ than just continuity.

    II. $if \int_{-\infty}^{\infty}|f(t)|dt \; < \; \infty \; then \; \mathcal{F}f(s) \to 0 \; as \; s \to \pm\infty.$

This is called the Riemann-Lebesgue lemma and it's more difficult to prove than showing simply that $\mathcal{F}f$is continuous. One might view the result as saying that $\mathcal{F}f(s)$ isat least tryingto be integrable. It's continuous and it tends to zero as $s \to \pm\infty$. Unfortunately, the factthat $\mathcal{F}f(s) \to 0$ does not imply that it's integrable (think of $sinc$, again). If we knew something, or couldinsist on something about the rateat which a signal or its transform tends to zero at $\pm\infty$ then perhapswe could push on further.

In the following we illustrate the, path , the way.To repeat, we want our theory to encompass the following three points:

1. The allowed signals include _'s, unit steps, ramps,$sines, cosines$, and all other standard signals that the world's economy depends on.
2. The Fourier transform and its inverse are defined for all of these signals.
3. Fourier inversion works.

Fiddling around with $L^1(\mathbf{R})$ or substitutes, putting extra conditions on jumps — all have been used. Thepath to success lies elsewhere. It is well marked and firmly established, but it involves a break with theclassical point of view. The outline of how all this is settled goes like this:

1. We single out a collection of functions s for which convergence of the Fourier integrals is assured,for which a function andits Fourier transform are both in s,and for which Fourier inversion works.
Furthermore, Parseval's identity holds:

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = \int_{-\infty}^{\infty} |\mathcal{F}f(s)|^2 \, ds.$$

This much isclassical; new ideas with new intentions, yes, but not new objects. Perhaps surprisinglyit's not so hard to find a suitable collection s, at least if one knows what one is looking for. Butwhat comes next is definitely not "classical". It had been first anticipated and used effectively in anearly form by O. Heaviside, developed, somewhat, and dismissed, mostly, soon after by less talentedpeople, then cultivated by and often associated with the work of P. Dirac, and finally defined by$L$. Schwartz.

2. s forms a class of test functionswhich, in turn, serve to define a larger class of generalized functions or distributions, called, for this class of test functions the tempereddistributions, T . Precisely because$s$ was chosen to be the ideal Fourier friendly space of classical signals, the tempered distributionsare likewise well suited for Fourier methods. The collection of tempered distributions includes, for

example, L^1 and L^2-functions (which can be wildly discontinuous), the $sinc$function, and complexexponentials (hence periodic functions). But it includes much more, like the delta functions andrelated objects.

3. The Fourier transform and its inverse will be defined so as to operate on these tempered distributions,and they operate to produce distributions of the same type. Thus the inverse Fourier transform canbe applied, and the Fourier inversion theorem holds in this setting.

4. In the case when a tempered distributions "comes from a function" — in a way we'll make precise— the Fourier transform reduces to the usual definition as an integral, when the integral makessense. However, tempered distributions are more general than functions, so we really will have donesomething new and we won't have lost anything in the process.

Our goal is to hit the relatively few main ideas in the outline above, suppressing the considerable massof details. In practical terms this will enable us to introduce delta functions and the like as tools forcomputation, and to feel a greater measure of confidence in the range of applicability of the formulas.

We're taking this path because it works, it's very interesting, and it's easy to compute with.

We'll touch on some other approaches to defining distributions and generalized Fourier transforms, butas far as we concerned they are the equivalent of vacuum tube technology.

Now we discuss The Right Functions for Fourier Transforms: Rapidly Decreasing

Functions. Mathematics progresses more by making intelligent definitions than by proving theorems. The hardestwork is often in formulating the fundamental concepts in the right way, a way that will then make thedeductions from those definitions (relatively) easy and natural. This can take awhile to sort out, and a subject might be reworked several times as it matures; when new discoveries are made and one seeswhere things end up, there's a tendency to go back and change the starting point so that the trip becomeseasier. Mathematicians may be more self-conscious about this process, but there are certainly examplesin engineering where close attention to the basic definitions has shaped a field — think of Shannon's workon Information Theory, for a particularly striking example.

Nevertheless, engineers, in particular, often find this tiresome, wanting to do something and not "just talkabout it": "Devices don't have hypotheses", as one of my colleagues put it. One can also have too muchof a good thing — too many trips back to the starting point to rewrite the rules can make it hard tofollow the game, especially if one has already played by the earlier rules. we sympathetic to both of thesecriticisms, and for our present work on the Fourier transform we'll try to steer a course that makes thedefinitions reasonable and lets us make steady forward progress.

In the following we study Smoothness and decay. To ask "how fast" $\mathcal{F}f(s)$ might tend to zero, depending on what additional assumptions we might makeabout the function $f(x)$ beyond integrability, will lead to our defining "rapidly decreasingfunctions",and this is the key. Integrability is too weak a condition on the signal f, but it does imply that $\mathcal{F}f(s)$ iscontinuous and tends to 0 $at \pm \infty$. What we're going to do is study the relationship between the smoothnessof a function — not just continuity, but how many times it can be differentiated — and the rate at whichits Fourier transform decays at infinity.

We'll always assume that $f(x)$ is absolutely integrable, and so has a Fourier transform. Let's suppose,more stringently, that

$$xf(x) \ is \ integrable, i.e., \int_{-\infty}^{\infty} |xf(x)|dx \ < \ \infty.$$

Then $xf(x)$ has a Fourier transform, and so does $-2\pi ixf(x)$ and its Fourier transform is

$$\mathcal{F}\big(-2\pi ixf(x)\big) \int_{-\infty}^{\infty} (-2\pi ix)e^{-2\pi isx} \ f(x)dx$$

$$= \int_{-\infty}^{\infty} \left(\frac{d}{ds}e^{-2\pi isx}\right) f(x)dx = \frac{d}{ds} \int_{-\infty}^{\infty} (e^{-2\pi isx})f(x)dx$$

(switching d/ds and the integral is justified by the integrability of $|x \ f(x)|$)

$$= \frac{d}{ds}(\mathcal{F}f)(s)$$

This says that the Fourier transform $\mathcal{F}f(s)$ is differentiable and that its derivative is $\mathcal{F}(-2\pi i\, xf(x))$. When $f(x)$ is merely integrable we know that $\mathcal{F}f(s)$ is merely continuous, but with the extra assumption on the integrability of $xf(x)$ we conclude that $(\mathcal{F}f)(s)$ is actually differentiable. (And its derivative is continuous. Why?)

For one more go-round in this direction, what if $x^2 f(x)$ is integrable? Then, by the same argument,

$$\mathcal{F}\big((-2\pi ix)^2 f(x)\big) = \int_{-\infty}^{\infty} (-2\pi ix)^2\, e^{-2\pi isx} f(x)\, dx$$

$$= \int_{-\infty}^{\infty} \left(\frac{d^2}{ds^2} e^{-2\pi isx}\right) f(x)\, dx = \frac{d^2}{ds^2}\int_{-\infty}^{\infty} (e^{-2\pi isx}) f(x)\, dx = \frac{d^2}{ds^2}(\mathcal{F}f)(s),$$

and we see that $\mathcal{F}f$ is twice differentiable. (And its second derivative is continuous.)

Clearly we can proceed like this, and as a somewhat imprecise headline we might then announce:

Faster decay of $f(x)$ at infinity leads to a greater smoothness of the Fourier transform.

Now let's take this in another direction, with an assumption on the smoothness of the signal. Suppose $f(x)$ is differentiable, that its derivative is integrable, and that $f(x) \to 0\ as\ x \to \pm\infty$. we've thrown in all the assumptions we need to justify the following calculation:

$$\mathcal{F}f(s) = \int_{-\infty}^{\infty} e^{-2\pi isx} f(x)\, dx$$

$$= \left[f(x)\frac{e^{-2\pi isx}}{-2\pi is}\right]_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} \frac{e^{-2\pi isx}}{-2\pi is} f'(x)dx$$
$$\qquad (integration\ by\ parts\ with\ u = f(x), \quad dv = e^{-2\pi isx}dx)$$
$$= \frac{1}{2\pi is}\int_{-\infty}^{\infty} e^{-2\pi isx} f'(x)dx\ (using\ f(x) \to 0\ as\ x \to \pm\infty)$$

$$= \frac{1}{2\pi is}(\mathcal{F}f')(s)$$

We then have

$$|\mathcal{F}f(s)| = \frac{1}{2\pi s}\left|(\mathcal{F}f')(s)\right| \le \frac{1}{2\pi s}\|f'\|_1.$$

The last inequality follows from the result: "The Fourier transform is bounded by the L^1-norm of thefunction". This says that $\mathcal{F}f(s)$ *tends to* 0 *at* $\pm \infty$ *like* $1/s$.(Remember that $\|f'\|_1$ is some fixed number here, independent of s.) Earlier we commented that if f is integrable then $\mathcal{F}f$ tends to 0 at $\pm\infty$, but here with the stronger assumptions we get a stronger conclusion, that $\mathcal{F}f$ tends to zero at acertain rate.

Let's go one step further in this direction. Suppose $f(x)$is twice differentiable, that its first and secondderivatives are integrable, and that $f(x)$and $f'(x)$ tend to 0 as

$x \to \pm \infty$. The same argument gives

$$\mathcal{F}f(s) = \int_{-\infty}^{\infty} e^{-2\pi i s x} f(x)\, dx$$

$= \frac{1}{2\pi i s} \int_{-\infty}^{\infty} e^{-2\pi i s x} f'(x) dx$ (picking up on where we were before)

$= \frac{1}{2\pi i s} \left(\left[f'(x) \frac{e^{-2\pi i s x}}{-2\pi i s} \right]_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} \frac{e^{-2\pi i s x}}{-2\pi i s} f''(x) dx \right)$

(integration by parts with $u = f'(x),\ dv = e^{-2\pi i s x} dx$)

$$\frac{1}{(2\pi i s)^2} \int_{-\infty}^{\infty} e^{-2\pi i s x} f''(x)\, dx \text{ (using } f'(x) \to 0 \text{ as x} \to \pm \infty)$$

$$= \frac{1}{(2\pi i s)^2} (\mathcal{F}f'')(s)$$

Thus

$$| \mathcal{F}f(s) | \le \frac{1}{(2\pi s)^2} \|f''\|_1$$

and we see that $\mathcal{F}f(s)$tends to 0 like $1/s^2$.

The headline:

Greater smoothness of $f(x)$, plus integrability, leads to faster decay of the Fourier transform at ∞.

Remark (3.1.1): (on the derivative formula for the Fourier transform)

In this work we rederived the derivative formula

$$\mathcal{F}f'(s) = 2\pi i s\ \mathcal{F}f(s)$$

which we've used before, but here we needed the assumption that

$f(x) \to 0$, whichwe didn't mentionbefore. What's up? With the technology we have available to us now, the derivation we gave, above, isthe correct derivation. That is, it proceedsvia integration by parts, and requires some assumption like

$f(x) \to 0\ as\ x \to \pm \infty$. That only works if Fourier inversion is known to hold. This was OK when therigor police were off duty, but not now, on this day of reckoning. when we develop a generalization of the Fourier transform.

We could go on as we did above, comparing the consequences of higher differentiability, integrability,smoothness and decay, bouncing back and forth between the function and its Fourier transform. The greatinsight in making use of these observations is that the simplest and most useful way to coordinate all these
phenomena is to allow for arbitrarily great smoothness and arbitrarily fast decay. We would like to haveboth phenomena in play. Here is the crucial definition.

definition (3.1.2): (Rapidly decreasing functions)
A function $f(x)$is said to be rapidly decreasing at $\pm \infty$ if
1. It is infinitely differentiable.
2. For all positive integers m and n,

$$|x^m \frac{d^n}{dx^n}| \to 0 \; as \; x \to \pm\infty$$

In words, any positive power of x times any order derivative of f tends to zero at infinity.
Note that m and n are independent in this definition. That is, we insist that, say, the 5th power of x timesthe 17th derivative of $f(x)$ tends to zero, and that the 100th power of x times the first derivative of $f(x)$ tends to zero; and whatever you want.
Are there any such functions? Any infinitely differentiable function that is ientically zero outside somefinite interval is one example, and we'll even write down a formula for one of these . Another example is$f(x) = e^{-x^2}$. You may already be familiar with the phrase "the exponential grows faster than any power of x", and likewise with the phrase "e^{-x^2} decays faster than any power of x." In fact, any derivative ofe^{-x^2}decays faster than any power of x as $x \to \pm\infty$, as you can check with L'Hopital's rule, for example.
We can express this exactly as in the definition:

$$|x^m \frac{d^n}{dx^n} e^{-x^2}| \to 0 \; as \; x \to \pm\infty$$

There are plenty of other rapidly decreasing functions. We also remark that if $f(x)$ is rapidly decreasingthen it is in $L^1(R)$ and in $L^2(R)$.

definition(3.1.3): (An alternative definition)
An equivalent definition for a function to be rapidly decreasing is to assumethat for any positive integers m and n there is a constant C_{mn} such that

$$|x^m \frac{d^n}{dx^n} e^{-x^2}| \leq C_{mn} \; as \; x \to \pm\infty$$

In words, the mth power of x times the nth derivative of$f$ remains bounded for all $m \; and \; n$, though theconstant will depend on which m and n we take. This
 Conditionimplies the "tends to zero" condition, above. Convince that, the key being that m and n are arbitrary and independent. We'll usethis second,

equivalent condition often, and it's a matter of taste which one takes as a definition.

Let us now praise famous men It was the French mathematician Laurent Schwartz who singled outthis relatively simple condition to use in the service of the Fourier transform. In his honor the set of rapidlydecreasing functions is usually denoted by S (a script S) and called the Schwartz class of functions.

Let's start to see why this was such a good idea.

1. The Fourier transform of a rapidly decreasing function is rapidly decreasing. Let $f(x)$bea function in S. We want to show that $\mathcal{F}f(s)$is also in S. The condition involves derivatives of $\mathcal{F}f$, sowhat comes in is the derivative formula for theFourier transform and the version of that formula for higherderivatives. As we've already seen

$$2\pi i s\ \mathcal{F}f(s) = \left(\mathcal{F}\frac{d}{dx}f\right)(s).$$

As we also noted,

$$\frac{d}{ds}\mathcal{F}f(s) = \mathcal{F}\left(-2\pi i x f(x)\right).$$

Because $f(x)$ is rapidly decreasing, the higher order versions of these formulas are valid; the derivationsrequire either integration by parts or differentiating under the integral sign, both of which are justified.

That is,

$$(2\pi i s)^n\ \mathcal{F}f(s) = \left(\mathcal{F}\frac{d^n}{dx^n}f\right)(s)$$

$$\frac{d^n}{ds^n}\mathcal{F}f(s) = \mathcal{F}\left((-2\pi i x)^n f(x)\right).$$

(We follow the convention that the zeroth order derivative leaves the function alone.)

Combining these formulas one can show, inductively, that for all nonnegative integers m and n,

$$\mathcal{F}\left(\frac{d^n}{dx^n}\left((-2\pi i x)^m f(x)\right)\right) = \frac{d^m}{ds^m}\mathcal{F}f(s).$$

Note how m and n enter in the two sides of the equation.

We use this last identity together with the estimate for the Fourier transform in terms of the L^1-norm ofthe function. Namely,

$$|s|^n\left|\frac{d^m}{ds^m}\mathcal{F}f(s)\right| = (2\pi)^{m-n}\left|\mathcal{F}\left(\frac{d^n}{dx^n}\left(x^m f(x)\right)\right)\right|$$

$$\leq (2\pi)^{m-n}\left\|\frac{d^n}{dx^n}\left(x^m f(x)\right)\right\|_1$$

The L^1-norm on the right hand side is finite because f is rapidly decreasing. Since the right hand sidedepends on $m\ and\ n$, we have shown that there is a constant $C_{mn}$with

$$\left| s^n \frac{d^m}{ds^m} \mathcal{F}f(s) \right| \leq C_{mn}$$

This implies that $\mathcal{F}f$ is rapidly decreasing. Done.

2. Fourier inversion works on S. We first establish the inversion theorem for a timelimited function in S. Suppose that $f(t)$ is smooth and for some T is identically zero for $|t| \geq T/2$, rather than just tendingto zero at $\pm\infty$. In this case we can eriodize $f(t)$ to get a smooth, periodic function of period T. Expandthe periodic function as a converging Fourier series. Then for $-T/2 \leq T/2$,

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{2\pi i n t}{T}}$$

$$= \sum_{n=-\infty}^{\infty} e^{\frac{2\pi i n t}{T}} \left(\frac{1}{T} \int_{-T/2}^{T/2} e^{\frac{-2\pi i n x}{T}} f(x) dx \right)$$

$$= \sum_{n=-\infty}^{\infty} e^{\frac{2\pi i n t}{T}} \left(\frac{1}{T} \int_{-\infty}^{\infty} e^{\frac{-2\pi i n x}{T}} f(x) dx \right) = \sum_{n=-\infty}^{\infty} e^{\frac{2\pi i n t}{T}} \mathcal{F}f\left(\frac{n}{T}\right) 1/T$$

Our intention is to let T get larger and larger. What we see is a Riemann sum for the integral

$$\int_{-\infty}^{\infty} e^{2\pi i s t} \mathcal{F}f(s) ds = \mathcal{F}^{-1}\mathcal{F}f(t),$$

and the Riemann sum converges to the integral because of the smoothness of f. (we have not slippedanything past you here, but we don't want to quote the precise results that make all this legitimate.) Thus

$$f(t) = \mathcal{F}^{-1}\mathcal{F}f(t),$$

and the Fourier inversion theorem is established for timelimited functions in S.

When f is not timelimited we use "windowing". The idea is to cut $f(t)$ off smoothly. The interestingthing in the present context— for theoretical rather than practical use — is to make the window so smooththat the "windowed" function is still in S.

We take a function $c(t)$ that is identically 1 for $-1/2 \leq t \leq 1/2$, that goes smoothly (infinitely differentiable)down to zero as t goes from 1/2 to 1 and from $-1/2$ to -1, and is then identically 0 for $t \geq 1$ and $t \leq -1$.This is a smoothed version of the rectangle function $\Pi(t)$; instead of cutting off sharply at $\pm 1/2$ we bringthe function smoothly down to zero. You can certainly imagine drawing such a function:
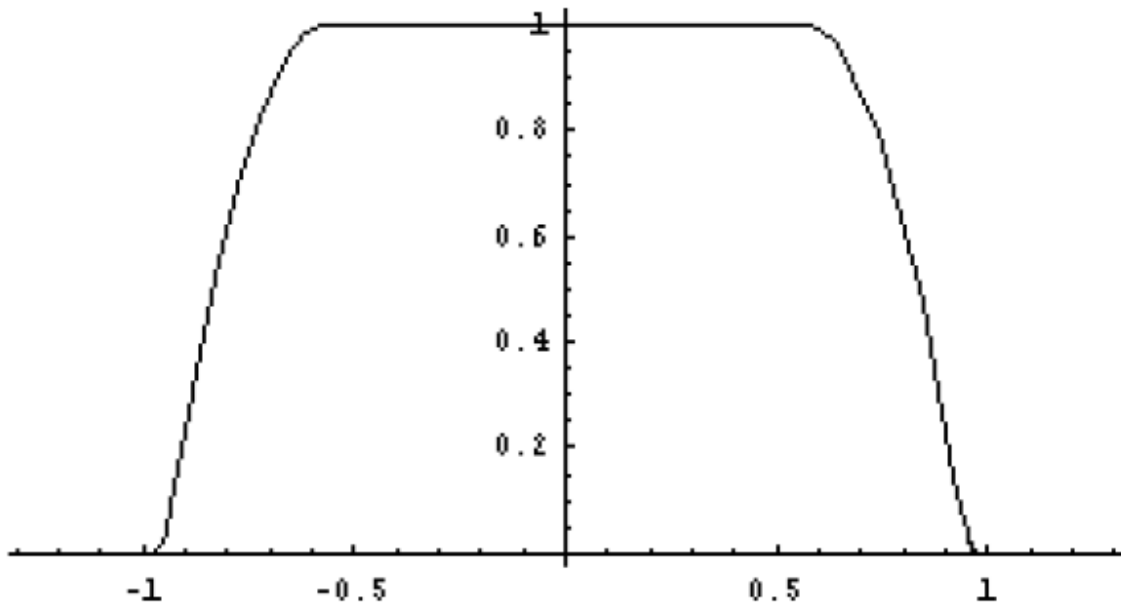
Figure (3.1)

Now scale $c(t)$ to $c_n(t) = c(t/n)$. That is, $c_n(t)$ is 1 for t between $-n/2$ and $n/2$, goes smoothly downto 0 between $\pm n/2$ and $\pm$ n and is then identically 0 for $|t| \geq n$. Next, the function $f_n(t) = c_n(t) \cdot f(t)$ isa timelimited function in S. Hence the earlier reasoning shows that the Fourier inversion theorem holdsfor f_n and $\mathcal{F}f_n$. The window eventually moves past every t, that is, $f_n(t) \to f$ (t) as
$n \to \infty$. Some
estimates based on the properties of the cut-off function .
3. Parseval holds in S. We'll actually derive a more general result than Parseval's identity, namely:
If $f(x)$ and $g(x)$ are complex valued functions in S then

$$\int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx = \int_{-\infty}^{\infty} \mathcal{F}f(s)\overline{\mathcal{F}g(s)}\,ds.$$

As a special case, if we take $f = g$ then $f(x)\overline{f(x)} = |f(x)|^2$ and the identity becomes

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\mathcal{F}f(s)|^2 ds.$$

To get the first result we'll use the fact that we can recover g from its Fourier transform via the inversiontheorem. That is,

$$g(x) = \int_{-\infty}^{\infty} \mathcal{F}g(s)e^{2\pi ist}\,ds.$$

The complex conjugate of the integral is the integral of the complex conjugate, hence

$$\overline{g(x)} = \int_{-\infty}^{\infty} \overline{\mathcal{F}g(s)}e^{-2\pi isx}\,ds.$$

The derivation is straightforward, using one of our favorite tricks of interchanging the order of integration:

$$\int_{-\infty}^{\infty} f(x)\,\overline{g(x)}\,dx = \int_{-\infty}^{\infty} f(x)\Big(\int_{-\infty}^{\infty} \overline{\mathcal{F}g(s)}e^{-2\pi isx}\,ds.\Big)dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\overline{\mathcal{F}g(s)}e^{-2\pi isx}\,ds\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)\overline{\mathcal{F}g(s)}e^{-2\pi isx}\,dx\,ds$$

$$= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} f(x)e^{-2\pi isx}\,dx\Big)\overline{\mathcal{F}g(s)}\,ds$$

$$= \int_{-\infty}^{\infty} \mathcal{F}f(s)\overline{\mathcal{F}g(s)}\,ds$$

All of this works perfectly — the initial appeal to the Fourier inversion theorem, switching the order ofintegration — if f and g are rapidly decreasing.

Now we discuss a very little on integrals. This discuss on integrals is not a short course on integration. It's here to provide a little. The star of this study here you go.

In the following we illustrate integrals are first defined for positive functions. In the general approach to integration (of realvaluedfunctions) you first set out to define the integral for nonnegative functions. Why? Because howevergeneral a theory you're constructing, an integral is going to be some kind of limit of sums and you'll want toknow when that kind of limit exists. If you work with positive (or at least nonnegative) functions then theissues for limits will be about how big the function gets, or about how big the sets are where the functionis or isn't big. You feel better able to analyze accumulations than to control conspiratorial cancellations.

So you first define your integral for functions $f(x)$ with $f(x) \geq 0$. This works fine. However, you knowfull well that your definition won't be too useful if you

can't extend it to functions which are both positiveand negative. Here's how you dothis. For any function $f(x)$you let $f^+(x)$be its positive part:
$$f^+(x) = max\{f(x), 0\}$$
Likewise, you let
$$f^-(x) = max\{f(x), 0\}$$
be its negative part. (Tricky: the "negative part" as you've defined it is actually a positive function; taking$-f(x)$ flips over the places where $f(x)$ is negative to be positive. You like that kind of thing.) Then

$$f = f^+ - f^-$$

while
$$|f| = f^+ + f^-.$$

You now say that f is integrable if both f^+ and $f^-$are integrable — a condition which makes sense sincef^+ and $f^-$are both nonnegative functions — and by definition you set

$$\int f = \int f^+ - \int f^-.$$

(For complex-valued functions you apply this to the real and imaginary parts.) You follow this approachfor integrating functions on a finite interval or on the whole real line. Moreover, according to this definition$|f|$ is integrable if f is because then

$$\int |f| = \int (f^+ + f^-) = \int f^+ + \int f^-$$

and f^+ and f^- are each integrable. It's also true, conversely, that if $|f|$is integrable then so is f.Youshow this by observing that
$$f^+ \leq |f| \qquad and\ f^- \leq |f|$$
and this implies that both f^+ and $f^-$are integrable.

You now know where the implication $\int_{-\infty}^{\infty} |f(t)|dt < \infty \Rightarrow \mathcal{F}f$exists comes from.

You get an easy inequality out of this development:
$$\left| \int f \right| \leq \int |f|.$$
In words, "the absolute value of the integral is at most the integral of the absolute value". And sure that'strue, because$\int f$.may involve cancellations of the positive and negative values of f while $\int |f|$won't have such cancellations. we don't shirk from a more formal argument:

$$\left| \int f \right| = \left| \int (f^+ - f^-) \right| = \left| \int f^+ - \int f^- \right|$$
$$\leq \left| \int f^+ \right| + |f^-| = \int f^+ + \int f^- \text{ (since } f^+ \text{ and } f^- \text{are both nonnegative)}$$
$$= \int (f^+ + f^-) = \int |f|.$$

You now know where the second inequality in

$$\left|\mathcal{F}f(s) - \mathcal{F}f(s')\right| = \left|\int_{-\infty}^{\infty} \left(e^{-2\pi i s t} - e^{-2\pi i t s'}\right) f(t)dt\right|$$

$$\leq \int_{-\infty}^{\infty} \left|e^{-2\pi i s t} - e^{-2\pi i t s'}\right| |f(t)|dt$$

comes from; this came up in showing that $\mathcal{F}f$ is continuous.

Now we illustrate *sinc* stinks What about the sinc function and trying to make sense of the following equation?

$$\mathcal{F}sinc(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} sinc\, t\, dt$$

According to the definitions you just gave, the *sinc* function is not integrable. In fact, the argument we gaveto show that

$$\int_{-\infty}^{\infty} |sinc\, t|dt = \infty$$

(the second argument) can be easily modified to show that both

$$\int_{-\infty}^{\infty} sinc^+t\, dt = \infty \ \ and \ \ \int_{-\infty}^{\infty} sinc^-t = \infty.$$

So if you wanted to write

$$\int_{-\infty}^{\infty} sinc\, t\, dt = \int_{-\infty}^{\infty} sinc^+t\, dt - \int_{-\infty}^{\infty} sinc^-t\, dt$$

we'd be faced with$1\infty - \infty$. Bad. The integral of sinc (and also the integral of $\mathcal{F}$sinc) has to be understoodas a limit,

$$\lim_{a\to-\infty, b\to\infty} \int_{-\infty}^{\infty} e^{-2\pi i s t} sinc\, t\, dt$$

Evaluating this is a classic of contour integration and the residue theorem.

In the following we study subtlety vs. cleverness. For the full mathematical theory of Fourier series and Fourier integrals oneneeds the Lebesgue integral. Lebesgue's approach to defining the integral allowsa wider class of functions to be integrated and it allows one to establish very general, very helpful resultsof the type "the limit of the integral is the integral of the limit", as in

$$f_n \to f \Rightarrow \lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(t)dt = \int_{-\infty}^{\infty} \lim_{n\to\infty} f_n(t)dt = \int_{-\infty}^{\infty} f(t)dt.$$

You probably do things like this routinely, and so do mathematicians, but it takes them a year or so ofgraduate school before they feel good about it. More on this in just a moment.

The definition of the Lebesgue integral is based on a study of the size, or measure, of the sets where afunction is big or small, and you don't wind up writing down the same kinds of "Riemann sums" youused in calculus to define the integral.

Nowtake note of the following quote of the mathematician T. K¨orner from his book Fourier Analysis:

  Mathematicians find it easier to understand and enjoy ideas which are clever rather than subtle.

  Measure theory is subtle rather than clever and so requires hard work to master.

Here'sone more thing:

The general result allowing one to pull a limit inside the integral sign is theLebesgue dominated convergencetheorem. It says: If f_n is a sequence of integrable functions that converges pointwise to a function f exceptpossibly on a set of measure 0, and if there is an integrable function g with $|f_n| \leq g$for all n (the"dominated" hypothesis) then f is integrable and

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} f_n(t)dt = \int_{-\infty}^{\infty} f(t)dt.$$

There's a variant of this that applies when the integrand depends on a parameter. It goes: If $f(x, t_0) = \lim_{t \to t_0} f(x, t)$ for all x, and if there is an integrable function g such that $|f(x,t)| \leq g(x)$ for all x then

$$\lim_{t \to t_0} \int_{-\infty}^{\infty} f(x,t)dt = \int_{-\infty}^{\infty} f(x,t_0)dx.$$

The situation described in this result comes up in many applications, and it's good to know that it holdsin great generality.

Integrals are not always just like sums. Here's one way they're different, and it's important to realizethis for our work on Fourier transforms.

For sums wehave the result that

$$\sum_n a_n convergesimpliesa_n \to 0.$$

We used this fact together with Parseval's identity for Fourier series to conclude that the Fourier coefficientstend to zero. You also all know the classic counterexample to the converse of the statement:

$1/n \to 0$ *but* $\sum_{n=1}^{\infty} \frac{1}{n}$diverge$s$.

For integrals, however, it is possible that

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

exists but $f(x)$ does not tend to zero at $\pm \infty$. Make $f(x)$ nonzero (make it equal to 1, if you want) onthinner and thinner intervals going out toward infinity. Then $f(x)$ doesn't decay to zero, but you can makethe intervals thin enough so that the integral converges.

How about this example? $\sum_{n=1}^{\infty} n\Pi(n^3(x-n))$

How shall we test for convergence of integrals? The answer depends on thecontext, and differentchoices are possible. Since the convergence of Fourier integrals is at stake, the important thing to measureis the size of a function "at infinity" — does it decay fast enough for the integrals to converge. Any kind of measuring requires a "standard", and for judging the decay (or growth) of a function the easiest andmost common standard is to measure using powers of x. The "ruler" based on powers of x reads:

$$\int_a^{\infty} \frac{dx}{x^p} \quad is \quad \begin{cases} infinite & if \ 0 < p \leq 1 \\ finite & if \ p > 1 \end{cases}$$

You can check this by direct integration. We take the lower limit a to be positive, but a particular value isirrelevant since the convergence or divergence of the integral depends on the decay near infinity. You canformulate the analogous statements for integrals $-\infty \ to - a$.

To measure the decay of a function $f(x) \ at \pm \ \infty$ we look at

$$\lim_{x\to\pm\infty} |x|^p|f(x)|$$

If, for some $p > 1$, this is bounded then $f(x)$ is integrable. If there is a $0 < p \leq 1$ for which the limit isunbounded, i.e., equals ∞, then $f(x)$ is not integrable.

Standards are good only if they're easy to use, and powers of x, together with the conditions on theirintegrals are easy to use. You can use these tests to show that every rapidly decreasing function is in both$L^1(R)$ and $L^2(R)$.

Section (3.2): Distributions and it's a physical analogy

We begin this section by studying distributions. Our program to extend the applicability of the Fourier transform has several steps. We took the first step last time:

We defined$s$, the collection of rapidly decreasing functions. In words, these are the infinitelydifferentiable functions whose derivatives decrease faster than any power of x at infinity. Thesefunctions have the properties that:

1. If $f(x)$ is in S then $\mathcal{F}f(s)$is in s.
2. If $f(x)$ is in S then $\mathcal{F}^{-1}\mathcal{F}f = f$.

We'll sometimes refer to the functions in s simply as Schwartz functions.

The next step is to use the functions in s to define a broad class of "generalized functions", or as we'll say,tempered distributions T , which will include s as

well as some nonintegrable functions, sine and cosine, δ functions, and much more, and for which the two properties, above, continue to hold.

We want to give a straightforward, no frills treatment of how to do this. There are two possible approaches.

1. Tempered distributions defined as limits of functions in s.

This is the "classical" (vacuum tube) way of defining generalized functions, and it pretty much appliesonly to the delta function, and constructions based on the delta function. This is an important enoughexample, however, to make the approachworth our while.

The other approach, the one we'll develop more fully, is:

2. Tempered distributions defined via operating on functions in s.

We also use a different terminology and say that tempered distributions are paired with functionsin s, returning a number for the pairing of a distribution with a Schwartz function.

In both cases it's fair to say that "distributions are what distributions do", in that fundamentally they aredefined by how they act on "genuine" functions, those in s. In the case of "distributions as limits", thenature of the action will be clear but the kind of objects that result from the limiting process is sort ofhazy. (That's the problem with this approach.) In the case of "distributions as operators" the nature of the objects is clear, but just how they are supposed to act is sort of hazy. (And that's the problem withthis approach, but it's less of a problem.) You may find the second approach conceptually more difficult,but removing the "take a limit"aspect from center stage really does result in a clearer and computationally

easier setup. The second approach is actually present in the first, but there it's cluttered up by framingthe discussion in terms of approximations and limits. Take your pick which point of view you prefer, butit's best if we comfortable with both.

In the following we discuss distributions as limitsThe first approach is to view generalized functions as some kind of limit of ordinary functions. Here we'll work with functions in s, but other functions can be used.

Let's consider the delta function as a typical and important example. we probably met δ as a mathematical,idealized impulse. we learned: "It's concentrated at the point zero, actually infinite at the pointzero, and it vanishes elsewhere." We learned to represent this graphically as a spike:
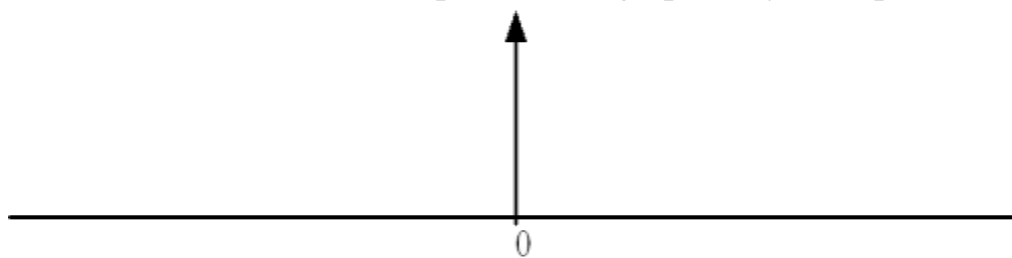

Figure (3.2)

we don't want to disabuse these ideas, or of the picture. We just want to refine thingssomewhat.

As an approximation to δ through functions in s one might consider the family of Gaussians

$$g(x,t) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}, \quad t > 0.$$

We remarked earlier that the Gaussians are rapidly decreasing functions.

Here's a plot of some functions in the family for t = 2, 1, 0.5, 0.1, 0.05 and 0.01. The smaller the valueof t, the more sharply peaked the function is at 0 (it's more and more "concentrated" there), while awayfrom 0 the functions are hugging the axis more and more closely. These are the properties we're trying tocapture, approximately.
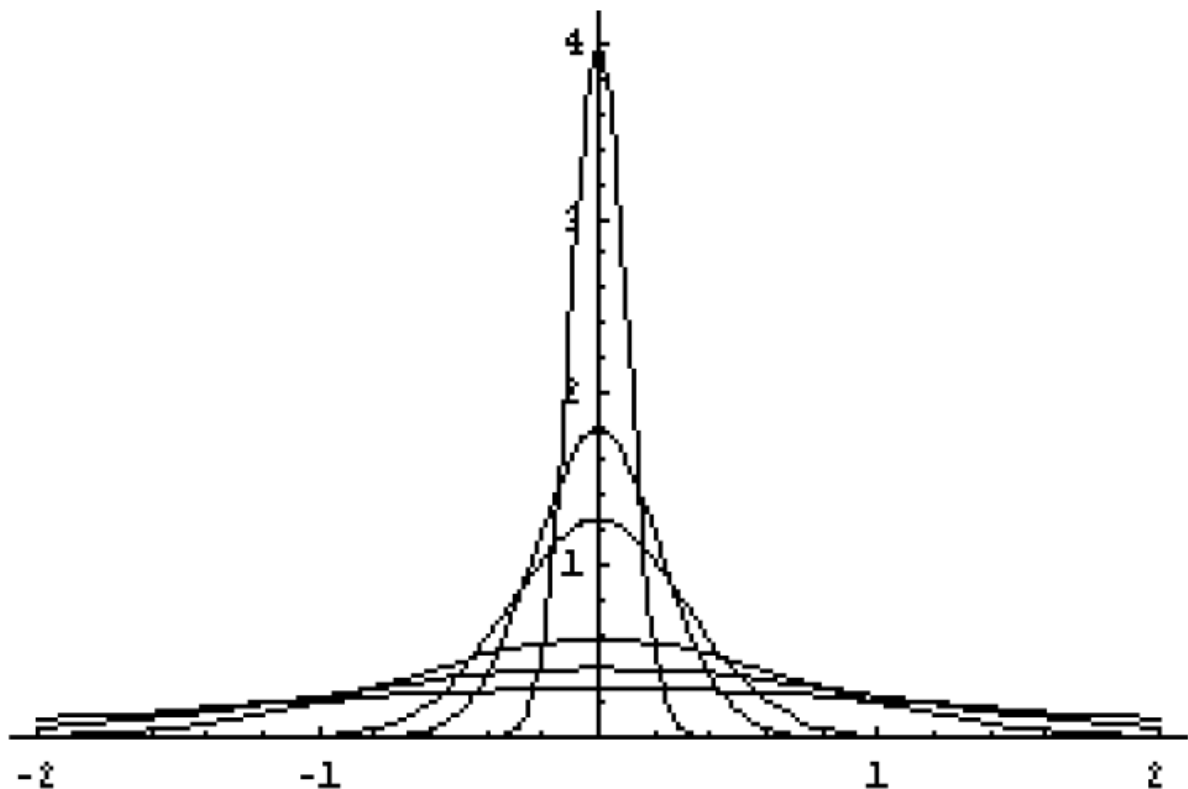
Figure (3.3)

As an idealization of a function concentrated at $x = 0, \delta$ should then be a limit

$$\delta(x) = \lim_{t \to 0} g(x,t).$$

This limit doesn't make sense as a pointwise statement — it doesn't define a function — but it begins tomake sense when one shows how the limit works operationally when "paired" with other functions. Thepairing, by definition, is by integration, and to anticipate the second approach to distributions, we'll write This as

$$\langle g(x,t), \varphi \rangle = \int_{-\infty}^{\infty} g(x,t)\,\varphi(x)dx.$$

(Don't think of this as an inner product. The angle bracket notation is just a good notation for pairing.)

The fundamental result — what it means for the $g(x,t)$ to be "concentrated at 0" as $t \to 0$ — is

$$\lim_{t\to 0} \int_{-\infty}^{\infty} g(x,t)\,\varphi(x)dx = \varphi(0).$$

Now, whereas we'll have a hard time making sense of $\lim_{t\to 0} g(x,t)$.alone,there's no trouble making senseof the limit of the integral, and, in fact, no trouble proving the statement just above. Do observe, however,that the statement: "The limit of the integral is the integral of the limit." is thus not true in this case.The limit of the integral makes sense but not the integral of the limit.

We can and will define the distribution δ by this result, and write

$$\langle \delta, \varphi \rangle = \lim_{t\to 0} \int_{-\infty}^{\infty} g(x,t)\,\varphi(x)dx = \varphi(0).$$

The Gaussians tend to ∞ $at\ x = 0\ as\ t \to 0,$ and that's why writing simply $\delta(x) = \lim_{t\to 0} g(x,t)$ doesn'tmake sense. One would have to say (and people do say, though I have a hard time with it) that the deltafunction has these properties:

1. $\delta(x) = 0\ for\ x \ne 0$
2. $\delta(0) = \infty$
3. $\int_{-\infty}^{\infty} \delta(x)dx = 1$

These reflect the corresponding (genuine) properties of the $g(x,t)$:

I. $\lim_{t\to 0} g(x,t) = 0\ \ if\ x \ne 0$
II. $\lim_{t\to 0} g(0,t) = \ \infty$
III. $\int_{-\infty}^{\infty} g(x,t)\,dx = 1$

The third property is our old friend, the second is clear from the formula, and you can begin to believe thefirst from the shape of the graphs. The first property is the flip side of "concentrated at a point", namelyto be zero away from the point where the function is concentrated.

The limiting process also works with convolution:

$$\lim_{t\to 0}(g * \varphi)(a) = \lim_{t\to 0} \int_{-\infty}^{\infty} g(a - x,t)\,\varphi(x)dx = \varphi(a).$$

This is written

$$(\delta * \varphi)(a) = \varphi(a)$$

as shorthand for the limiting process that got us there, and the notation is then pushed so far as to writethe delta function itself under the integral, as in

$$(\delta * \varphi)(a) = \int_{-\infty}^{\infty} \delta(a - x)\,\varphi(x)dx = \varphi(a).$$

The equation

$$(\delta * \varphi)(a) = \varphi(a)$$

completes the analogy: "δ is to 1 as convolution is to multiplication".

Why concentrate? Why would one want a function concentrated at a point in the first place? We'llcertainly have plenty of applications of delta functions very shortly. Heaviside used δ (without the notation) in hisapplications and reworking of Maxwell's theory of electromagnetism. The symbolism, and the three defining properties of δ listed above, were introduced later by P. Dirac in theservice of calculations in quantum mechanics. Because of Dirac's work, $\delta$is often referred to as the "Diracδ function".

For the present, let's take at the heat equation and how the delta function comes in there.

We're perfectly set up for that.

We have seen the family of Gaussians

$$g(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}, t > 0$$

Before. They arose in solving the heat equation for an "infinite rod". Recall that the temperature $u(x,t)$ at a point x and time t satisfies the partial differential equation

$$u_t = \frac{1}{2}u_{xx}.$$

When an infinite rod (the real line, in other words) is given an initial temperature $f(x)$ then $u(x,t)$ is givenby the convolution with $g(x,t)$:

$$u(x,t) = g(x,t) * f(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} * f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}} f(y)dy$$

One thing we didn't say at the time, knowing that this day would come, is how one recovers the initialtemperature $f(x)$ from this formula. The initial temperature is at t = 0, so this evidently requires that wetake the limit:

$$\lim_{t \to 0^+} u(x,t) = \lim_{t \to 0^+} g(x,t) * f(x) = (\delta * f)(x) = f(x).$$

Out pops the initial temperature.

Now we illustrate distributions as linear functional. Farewell to vacuum tubes The approach to distributions we've just followed, illustrated by defining δ,can be very helpful in particular cases and where there's a natural desire to have

everything look as"classical" as possible. Still and all, we maintain that adopting this approach wholesale to defining andworking with distributions is using technology from a bygone era. we haven't yet defined the collection oftempered distributions T which is supposed to be the answer to all our Fourier prayers, and we don't know

how to do it from a purely "distributions as limits" point of view. It's time to transistorize.

In the preceding discussion we did wind up by considering a distribution, at least δ, in terms of how it acts

when paired with a Schwartz function. We wrote

$$(\delta * \varphi) = \varphi(0)$$

as shorthand for the result of taking the limit of the pairing

$$\langle g(x,t), \varphi(x) \rangle = \int_{-\infty}^{\infty} g(x,t)\, \varphi(x) dx.$$

The second approach to defining distributions takes this idea — "the outcome" of a distribution actingon a test function — as a starting point rather than as a conclusion. The question to ask is what aspectsof "outcome", as present in the approach via limits, do we try to capture and incorporate in the basicdefinition?

Mathematical functions defined on **R**, "live at points", to use the hip phrase. That is, you plug in aparticular point from **R**, the domain of the function, and you get a particular value in the range, as forinstance in the simple case when the function is given by an algebraic expression and you plug values intothe expression. Generalized functions — distributions— do not live at points. The domain of a generalizedfunction is not a set of numbers. The value of a generalized function is not determined by plugging in anumber from **R** and determining a corresponding number. Rather, a particular value of a distribution isdetermined by how it "operates" on a particular test function. The domain of a generalized function is a

set of test functions. As they say in Computer Science, helpfully:

• You pass a distribution a test function and it returns a number.

That's not so outlandish. There are all sorts of operations you've run across that take a signal as anargument and return a number. The terminology of "distributions" and "test functions", from the dawn ofthe subject, is even supposed to be some kind of desperate appeal to physical reality to make this reworking

of the earlier approaches more appealing and less "abstract".

Having come this far, but still looking backward a little, recall that we asked which properties of a pairing— integration, as we wrote it in a particular case in

the first approach — do we want to subsume in thegeneral definition. To get all we need, we need remarkably little. Here's the definition:

Definition(3.2.3): (Tempered distributions)

A tempered distribution T is a complex-valued continuous linear functional on the collection S of Schwartz functions (called test functions). We denote the collection of all tempereddistributions by T.

That's the complete definition, but we can unpack it a bit:

1. If φ is in S then $T(\varphi)$ is a complex number. (You pass a distribution aSchwartz function, it returnsa complex number.)

We often write this action of T on φ as $(T * \varphi)$and say that T is paired with φ. (This terminologyand notation are conventions, not commandments.)

2. A tempered distribution is linear operating on test functions:
$$\mathrm{T}(\alpha_1\varphi_1 + \alpha_2\varphi_2) = \alpha_1\mathrm{T}(\varphi_1) + \alpha_2\mathrm{T}(\varphi_2)$$
or, in the other notation,
$$\langle \mathrm{T}, \alpha_1\varphi_1 + \alpha_2\varphi_2 \rangle = \alpha_1\langle \mathrm{T}, \varphi_1 \rangle + \alpha_2\langle \mathrm{T}, \varphi_2 \rangle,$$
for test functions $\varphi_1, \varphi_2$and complex numbers α_1, α_2.

3. A tempered distribution is continuous: if φ_n is a sequence of test functions in S with $\varphi_n \to \varphi$ in S then

$T(\varphi_n) \to T(\varphi)$, also written $\langle \mathrm{T}, \varphi_n \rangle \to \langle \mathrm{T}, \varphi \rangle$.

Also note that two tempered distributions $T_1 \; and \; T_2$ are equal if they agree on all test functions:

$T_1 = T_2$if $T_1(\varphi) = T_2(\varphi)$ $(\langle \mathrm{T}_1, \varphi \rangle = \langle \mathrm{T}_2, \varphi \rangle)$ for all $\varphi \; in \; S$.

This isn't part of the definition, it's just useful to write down.

There's a catch There is one hard part in the definition, namely, what it means for a sequence of testfunctions in S to converge in S. To say that$\varphi_n \to \varphi$ in S is to control the convergence of φ_n togetherwith all its derivatives. We won't enter into this, and it won't be an issue for us. If you look in standardmathematics books on the theory of distributions you will find long, difficult discussions of the appropriatetopologies on spaces of functions that must be used to talk about convergence. And you will be discouragedfrom going any further.

It's another question to ask why continuity is included in the definition. Let me just say that this isimportant when one considers limits of distributions and approximations to distributions.

Other classes of distributions This settles the question of what a tempered distribution is: it's acontinuous linear functional on S. For those who know the terminology, T is the dual space of the spaceS. In general, the dual space to a vector space is the set of continuous linear functionals on the vector space,the catch being to define continuity appropriately. From this point of view one can imagine defining typesof distributions other than the tempered distributions.

They arise by taking the dual spaces of collectionsof test functions other than S. Though we'll state things for tempered distributions, most general facts(those not pertaining to the Fourier transform, yet to come) also hold for other types of distributions.

In the following we discuss two important examples of distributions. Let us now understand:

1. How T somehow includes the functions we'd like it to include for the purposes of extending theFourier transform.

2. How δ fits into this new scheme.

The first item is a general construction and the second is an example of a specific distribution defined inthis new way.

How functions determine tempered distributions, and why the tempered distributions includethe functions we want. Suppose $f(x)$ is a function for which

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

exists for all Schwartz functions $\varphi(x)$. This is not asking too much, considering that Schwartz functionsdecrease so rapidly that they're plenty likely to make a product $f(x)\varphi(x)$integrable.

In this case the function $f(x)$ determines ("defines" or "induces" or "corresponds to"— pick your preferreddescriptive phrase) a tempered distribution $T_f$by means of the formula

$$T_f(\varphi) = \int_{-\infty}^{\infty} f(x)\varphi(x)dx$$

In words, T_f acts on a test function φ by integration of φ against f. Alternatively, we say that the function$f$ determines a distribution T_f through the pairing

$\langle T_f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)dx$, φ a test function.

This is just what we considered in the earlier approach that led to δ, pairing Gaussians with a Schwartzfunction. In the present terminology we would say that the Gaussian $g(x,t)$ determines a distribution $T_g$according to the formula

$$\langle T_g, \varphi \rangle = \int_{-\infty}^{\infty} g(x,t)\varphi(x)dx.$$

Let's check that the pairing $\langle T_f, \varphi \rangle$ meets the standard of the definition of a distribution. The pairing islinear because integration is linear:

$$\langle T_f, \alpha_1\varphi_1 + \alpha_2\varphi_2 \rangle = \int_{-\infty}^{\infty} f(x)\big(\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x)\big)dx$$

$$=\int_{-\infty}^{\infty} f(x)\alpha_1\varphi_1(x)dx + \int_{-\infty}^{\infty} f(x)\alpha_2\varphi_2(x)dx$$
$$=\alpha_1\langle T_f, \varphi_1\rangle + \alpha_2\langle T_f, \varphi_2\rangle$$

What about continuity? We have to take a sequence of Schwartz functions φ_n converging to a Schwartz function φ and consider the limit

$$\lim_{n\to\infty} \langle T_f, \varphi_n\rangle = \lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\,\varphi_n(x)dx.$$

Again, we haven't said anything precisely about the meaning of $\varphi_n \to \varphi$, but the standard results on taking the limit inside the integral will apply in this case and allow us to conclude that

$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\,\varphi_n(x)dx .= \int_{-\infty}^{\infty} f(x)\varphi(x)dx$

i.e., that

$$\lim_{n\to\infty} \langle T_f, \varphi_n\rangle = \langle T_f, \varphi\rangle.$$

This is continuity.

Using a function $f(x)$ to determine a distribution T_f this way is a very common way of constructing distributions. We will use it frequently. Now, you might ask whether different functions can give rise to the same distribution. That is, if $T f_1 = T f_2$ as distributions, then must we have $f_1(x) = f_2(x)$ (x)? Yes, fortunately, for if $T f_1 = T f_2$ then for all test functions $\varphi(x)$ we have

$$\int_{-\infty}^{\infty} f_1\,\varphi(x)dx = \int_{-\infty}^{\infty} f_2\varphi(x)dx$$

hence

$$\int_{-\infty}^{\infty} (f_1(x) - f_2(x))\varphi(x)dx = 0$$

Since this holds for all test functions $\varphi(x)$ we can conclude that $f_1(x =)f_2(x)$. Because a function $f(x)$ determines a unique distribution, it's natural to "identify" the function $f(x)$ with the corresponding distribution T_f. Sometimes we then write just f for the corresponding distribution rather than writing T_f, and we write the pairing as

$$\langle f, \varphi\rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)dx.$$

rather than as $\langle T_f, \varphi\rangle$.

• It is in this sense — identifying a function f with the distribution T_f it determines — that a class of distributions "contains" classical functions. Let's look at some examples.

Examples (3.2.4):

The *sinc* function defines a tempered distribution, because, though *Sinc* is not integrable,$(sincx) \varphi(x)$ is integrable for any Schwartz function $\varphi(x)$.Remember that a Schwartz function $\varphi(x)$diesoff faster than any power of x and that's more than enough to pull sinc down rapidly enough at $\pm\infty$tomake the integral exist. we not going to prove this but we have no qualms asserting it. For example, here'sa plot of e^{-x^2} times the sinc function on the interval
$$-3.5 \leq x \leq 3.5:$$
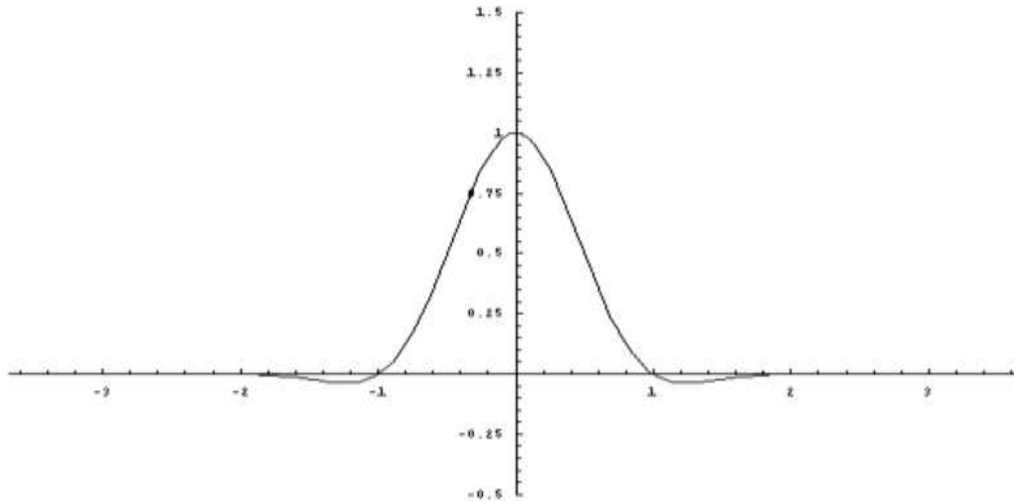
Figure(3.4)

For the same reason any complex exponential, and also sine and cosine, define tempered distributions.

Here's a plot of e^{-x^2} times $cos2\pi t$ on the range $-3.5 \leq x \leq 3.5$:
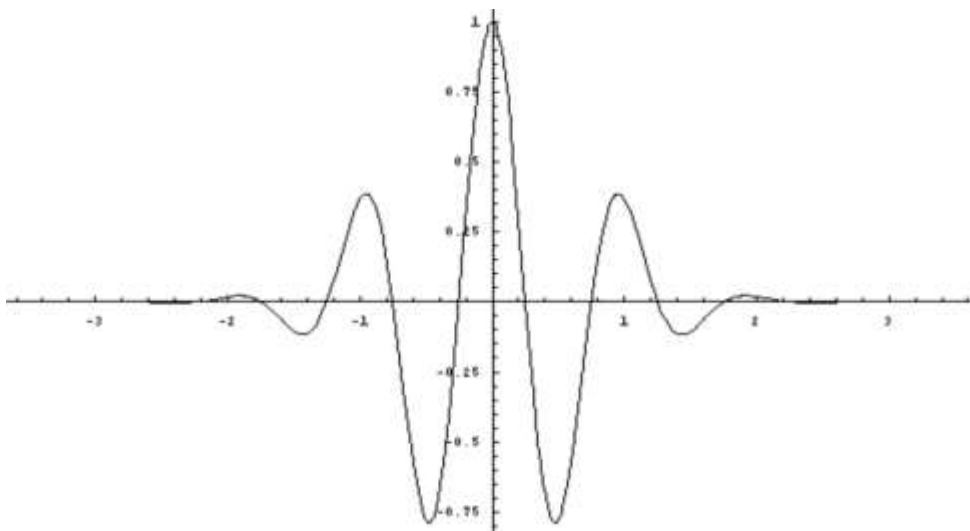
Figure(3.5)

Take two more examples, the Heaviside unit step $H(x)$and the unit ramp $u(x)$:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \qquad\qquad u(x) = \begin{cases} 0 & x \leq 0 \\ x & x \geq 0 \end{cases}$$

Neither function is integrable; indeed, $u(x)$even tends to ∞ as x $\to$ ∞, but it does so only to the first power (exactly) of x. Multiplying by a Schwartz function brings $H(x)$and $u(x)$down, and they eachdetermine tempered distributions. Here are plots of e^{-x^2} times $H(x)$and $u(x)$,respectively:
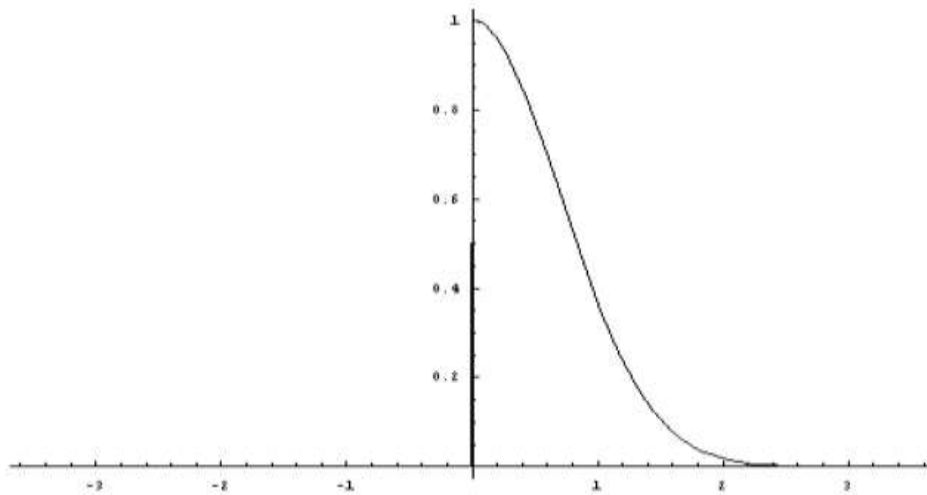
Figure(3.6)
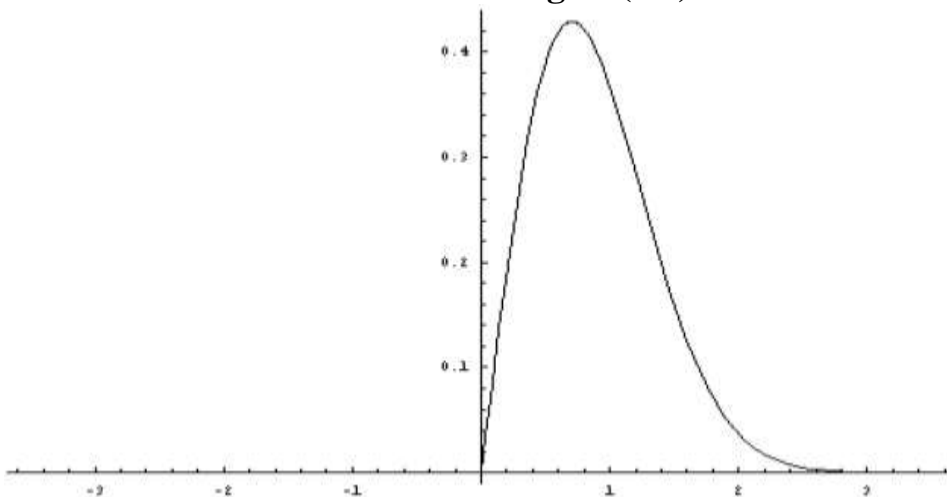
Figure(3.7)

The upshot is that the $sinc,$ complex exponentials, the unit step, the unit ramp, and many others, canall be considered to be tempered distributions. This is a good thing, because we're aiming to define theFourier transform of a tempered distribution, and we want to be able to apply it to the signals societyneeds. (We'll also get back to our good old formula $\mathcal{F}$ $sinc$ $=$ Π, and all will be right with the world.)

Do all tempered distributions "come from functions" in this way? δ does not come from a function in the way we'vejust described (or in any way). This adds to the feeling that we really have defined something new, that"generalized functions" include many (classical) functions but go beyond the classical functions.

Two final points. As we've just remarked, not every distribution comes from a function and so the natureof the pairing of a given tempered distribution T with a Schwartz function φ is unspecified, so to speak.

By that we mean, don't think that $\langle T, \varphi \rangle$is an integral, as in

$$\langle T, \varphi \rangle = \int_{-\infty}^{\infty} T(x)\varphi(x)dx$$

for any old tempered distribution T. The pairing is an integral when the distribution comes from afunction, but there's more to tempered distributions than that.

Finally a note of caution. Not every function determines a tempered distribution. For example e^{-x^2} doesn't. It doesn't because $e^{-x^2}$is a Schwartz function and

$$\int_{-\infty}^{\infty} e^{x^2} e^{-x^2} dx = \int_{-\infty}^{\infty} 1 \, dx = \infty.$$

δ as a tempered distribution The limiting approach to the delta function culminated with our writing

$$\langle \delta, \varphi \rangle = \varphi(0)$$

as the result of$\lim_{t \to \infty} \int_{-\infty}^{\infty} g(x,t) \, \varphi(x)dx = \varphi(0)$.

Now with our second approach, tempered distributions as linear functionals on S, we can simply define the tempered distribution δ by how it should operate on a function φ in S so as to achieve this outcome,and obviously what we want is $\delta(\varphi) = \varphi(0)$, or in the bracket notation $\langle \delta, \varphi \rangle = \varphi(0)$;
you pass δ a test function and it returns the value of the test function at 0.

Let's check the definition. For linearity,

$$\langle \delta, \varphi_1 + \varphi_2 \rangle = \varphi_1(0) + \varphi_2(0) = \langle \delta, \varphi_1 \rangle + \langle \delta, \varphi_2 \rangle$$
$$\langle \delta, \alpha\varphi \rangle = \alpha\varphi(0) = \alpha\langle \delta, \varphi \rangle$$

For continuity, if$\varphi_n(x) \to \varphi(0)$then in particular $\varphi_n(0) \to \varphi(0)$ and so

$$\langle \delta, \varphi_n \rangle = \varphi_n(0) \to \varphi(0) = \langle \delta, \varphi \rangle$$

So the mysterious δ, clouded in controversy by statements like

$$\delta(x) = 0 \, for \, x \neq 0$$
$$\delta(0) = \infty$$
$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

now emerges as the simplest possible nontrivial tempered distribution — it's just the functional describedin words by "evaluate at 0".

There was a second identity that we had from the "δ as limit" development, namely

$$(\delta_a * \varphi) = \varphi(a).$$

as a result of

$$\lim_{t \to 0} \int_{-\infty}^{\infty} g(a - x, t)\, \varphi(\mathrm{x})\mathrm{dx} = \varphi(\mathrm{a})$$

we define a tempered distribution δ_a (the δ function based at a) by the formula

$$\langle \delta_a, \varphi \rangle = \varphi(\mathrm{a}).$$

In words, you pass δ_a a test function and it returns the value of the test function at a. we won't check that δ_a satisfies the definition — it's the same argument as for δ.

δ and δ_a are two different distributions ($for\ a\ a\ =\ 0$). Classically, if that word makes sense here, one would write δ_a as $\delta(x - a)$, just a shifted δ. We'll get to that, and use that notation too, but a bit later. As tempered distributions, δ and δ_a are defined to have the property we want them to have. It's air tight — no muss, no fuss. That's δ. That's δ_a.

Would we have come upon this simple, direct definition without having gone through the "distributions aslimits" approach? Would we have the transistor without first having vacuum tubes? Perhaps so, perhapsnot. That first approach via limits provided the basic insights that allowed people, Schwartz in particular,to reinvent the theory of distributions based on linear functionals as we have done here (as he did).

Now we illustrate other types of distributions. We have already seen that the functions in S work well for Fourier transforms. However, S isn't the only possible collection of test functions and τ isn't the only possiblecollection of distributions.

Another useful set of test functions are the smooth functions that are timelimited,

That is, we let C be the set of infinitely differentiable functions which areidentically zero beyond apoint:

$\varphi(x)\ is\ in\ C\ if\ \varphi(x)$ has derivatives of all orders and if $\varphi(x)\ =\ 0$ for $|x|\ \geq x_0$(where x_0 can depend on φ).

The mathematical terminology for such a function is that it has compact support. The support of a functionis the complement of the largest set where the function is identically zero. (The letter C is supposed toconnote "compact".)

The continuous linear functionals on$C$ also form a collection of distributions, denoted by ⍰. In fact, whenmost people use the term "distribution" (without the adjective tempered) they are usually thinking of anelement of ⍰. We use the same notation as before for the pairing: $\langle \mathrm{T}, \varphi \rangle$for T in ⍰ and φin C.

δ and δ_a belong to ⍰as well as to τ , and the definition is the same:

$\langle \delta, \varphi \rangle = \varphi(0)$ and $\langle \delta_a, \varphi \rangle = \varphi(a)$.

It's the same δ. It's not a new distribution, it's only operating on a different class of test functions.

$\mathcal{D}$ is a bigger collection of distributions than τ because C is a smaller collection of test functions than S. The latter point should be clear to you: To say that $\varphi(x)$ is smooth and vanishes identically outside some interval is a stronger condition than requiring merely that it decays at infinity (albeit faster than any power of x). Thus if $\varphi(x)$ *is in C* then it's also in S. Why is $\mathcal{D}$ bigger than ? Since C is contained in

S, a continuous linear functional on S is automatically a continuous linear functional on C. That is, τ is contained in $\mathcal{D}$.

Just as we did for , we say that a function $f(x)$ determines a distribution in $\mathcal{D}$ if

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

exists for all test functions φ *in C*. As before, we write T_f for the distribution induced by a function f, and the pairing as

$$\langle T_f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)dx.$$

As before, a function determines a unique distribution in this way, so we identify f with T_f and write the pairing as

$$\langle f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)dx.$$

It's easier to satisfy the integrability condition for C than for S because multiplying $f(x)$ by a function in C kills it off completely outside some interval, rather than just bringing it smoothly down to zero at infinity as would happen When

multiplying by a function in S. This is another reason why $\mathcal{D}$ is a bigger class of distributions than τ — more functions determine distributions. For example, we observed that the function e^{x^2} doesn't determine a tempered distribution, but it does determine an element of $\mathcal{D}$.

Now we study a Physical Analogy for distributions. Think of heat distributed over a region in space. A number associated with heat is temperature, and we want to measure the temperature at a point using a thermometer. But does it really make sense to ask for the temperature "at a point"? What kind of test instrument could possibly measure the temperature at a point? What makes more sense is that a thermometer registers some overall value of the temperature near a point.

That is, the temperature is whatever the thermometer says it is, and is determined by a pairing of the heat (the distribution) with the thermometer (a test function or test device). The more "concentrated" the thermometer (the more

sharply peaked the test function) the more accurate the measurement, meaningthe closer the reading is to being the temperature "at a point".

A pairing of a test function with the heat is somehow supposed to model how the thermometer respondsto the distribution of heat. One particular way to model this is to say that if f is the heat and φ is thetest function, then the reading on the thermometer is

$$\int f(x)\varphi(x)dx$$

an integrated, average temperature. we've left limits off the integral to suggest that it is taken over someregion of space where the heat is distributed.

Such measurements (temperature or other sorts of physical measurements) are supposed to obey laws ofsuperposition (linearity) and the like, which, in this model case, translates to

$$\int f(x)(\alpha_1\varphi_1(x) + \alpha_2\varphi_2(x))dx = \alpha_1 \int f(x)(\varphi_1(x)\, dx + \alpha_2 \int f(x)\varphi_2(x)dx$$

for test functions φ_1 and φ_2. That's why we incorporate linearity into the definition of distributions. Withenough wishful thinking you can pass from this motivation to the general definition. Sure you can.

In the following we discuss limits of distributionsThere's a very useful general result that allows us to define distributions by means of limits. The statement goes:

> Suppose that T_n is a sequence of tempered distributions and that $\langle T_n, \varphi \rangle$
> (asequence of numbers)converges for every Schwartz function φ. Then T_n
> converges to a tempered distribution Tand
> $$\langle T, \varphi \rangle = \lim_{n \to \infty} \langle T_n, \varphi \rangle$$

Briefly, distributions can be defined by taking limits of sequences of distributions, and the result says thatif the parings converge then the distributions converge. This is by no means a trivial fact, the key issuebeing the proper notion of convergence of distributions, and that's hard. We'll have to be content with the
statement and let it go at that.

You might not spot it from the statement, but one practical consequence of this result is that if differentconverging sequences have the same effect on test functions then they must be converging to the samedistribution. More precisely, if $\lim_{n \to \infty} \langle S_n, \varphi \rangle$and $\lim_{n \to \infty} \langle T_n, \varphi \rangle$ both exist and are equal for every test function φ then S_n and T_n both converge to the same distribution. That's certainly possible — differentsequences can have the same limit, after all.

To illustrate just why this is helpful to know, let's consider different ways of approximating δ.

Now we illustrate other approximating Sequences for . Earlier we used a family

of Gaussians to approach δ, but there are many other ways we could try to approximate this characteristic behavior of δ in the limit. For example, take the family of scaled Π functions

$$R_\varepsilon(x) = \frac{1}{\varepsilon}\Pi_\varepsilon(x) = \frac{1}{\varepsilon}\Pi\left(\frac{x}{\varepsilon}\right) = \begin{cases} \dfrac{1}{\varepsilon} & |x| < \dfrac{\varepsilon}{2} \\ 0 & |x| \geq \dfrac{\varepsilon}{2} \end{cases}$$

where ε is a positive constant. Here's a plot of $R_\varepsilon(x)$ for $\varepsilon = 2, 1, 0.5, 0.1$, some of the same values weused for the parameter in the family of Gaussians.
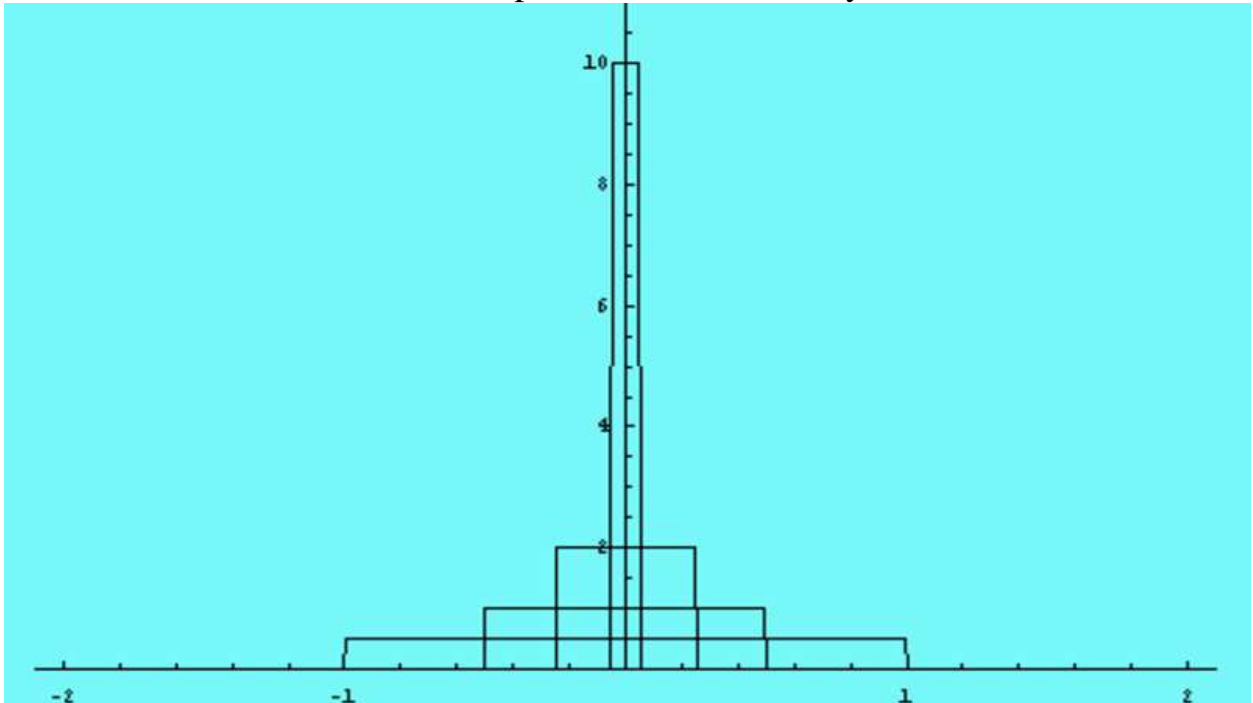
Figure (3.8)

What happens if we integrate $R_\varepsilon(x)$against a test function $\varphi(x)$? The function $\varphi(x)$ could be a Schwartzfunction, if we wanted to stay within the class of tempered distributions, or an element of C. In fact, allthat we require is that $\varphi(x)$ is smooth near the origin so that we can use a Taylor approximation (and wecould get away with less than that). We write

$$\langle R_\varepsilon, \varphi \rangle = \int_{-\infty}^{\infty} R_\varepsilon(x)\varphi(x)dx = \frac{1}{\varepsilon}\int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} \varphi(x)dx$$

$$= \frac{1}{\varepsilon}\int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} (\varphi(0) + \varphi''(0)x + 0(x^2)dx = \varphi(0) + \int_{-\frac{\varepsilon}{2}}^{\frac{\varepsilon}{2}} 0(x^2)dx = \varphi(0) + 0(\varepsilon^2)$$

If we let $\varepsilon \to 0$ we obtain

$$\lim_{\varepsilon \to 0}\langle R_\varepsilon, \varphi \rangle = \varphi(0).$$

113

In the limit, the result of pairing the R_ε with a test function is the same as pairing a Gaussian with a testfunction:

$$\lim_{\varepsilon \to 0} \langle R_\varepsilon, \varphi \rangle = \varphi(0) = \lim_{t \to 0} \langle g(x,t), \varphi(x) \rangle.$$

Thus the distributions defined by R_ε and by $g(x,t)$ each converge and to the same distribution, namely δ.

A general way to get to δ There's a general, flexible and simple approach to getting to δ by a limit.

It can be useful to know this if one model approximation might be preferred to another in a particularcomputation or application. Start with a function $f(x)$ having

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

and form

$$f_p(x) = pf(px), \quad p > 0.$$

Then one has

$$f_p \to \delta .$$

How does f_p compare with f? As p increases, the scaled function $f(px)$ concentrates near x = 0, thatis, the graph is squeezed in the horizontal direction. Multiplying by p to form p $f(px)$ then stretches thevalues in the vertical direction. Nevertheless

$$\int_{-\infty}^{\infty} f_p(x)dx = 1$$

as we see by making the change of variable $u = px$.

To show that f_p converges to δ, we pair $f_p(x)$ with a test function $\varphi(x)$ via integration and show

$$\lim_{p \to \infty} \int_{-\infty}^{\infty} f_p(x) \varphi(x)dx = \varphi(0) = \langle \delta, \varphi \rangle .$$

There is a nice argument to show this. Write

$$\int_{-\infty}^{\infty} f_p(x)\varphi(x)dx = \int_{-\infty}^{\infty} f_p(x)(\varphi(x) - \varphi(0) + \varphi(0))dx$$

$$\int_{-\infty}^{\infty} f_p(x)(\varphi(x) - \varphi(0))dx + \varphi(0) \int_{-\infty}^{\infty} f_p(x)dx$$

$\int_{-\infty}^{\infty} f_p(x)(\varphi(x) - \varphi(0))dx + \varphi(0),$

$$\int\limits_{-\infty}^{\infty} f_p(x)\big(\varphi(x/p) - \varphi(0)\big)dx + \varphi(0),$$

where we have used that the integral of f_p is 1 and have made a change of variable in the last integral.

The object now is to show that the integral of $f(x)\big(\varphi(x/p) - \varphi(0)\big)$ goes to zero as $p \to \infty$. There aretwo parts to this. Since the integral of $f(x)\big(\varphi(x/p) - \varphi(0)\big)$ is finite, the tails at $\pm\infty$ are arbitrarily small,meaning, more formally, that for any $\varepsilon > 0$ there is an $a > 0$ such that

$$\left|\int\limits_{a}^{\infty} f(x)\left(\varphi\left(\frac{x}{p}\right) - \varphi(0)\right)dx\right| + \left|\int\limits_{-\infty}^{-a} f(x)\left(\varphi\left(\frac{x}{p}\right) - \varphi(0)\right)dx\right| < \varepsilon.$$

This didn't involve letting p tend to ∞; that comes in now. Fix a as above. It remains to work with theintegral

$$\int\limits_{-a}^{a} f(x)\left(\varphi\left(\frac{x}{p}\right) - \varphi(0)\right)dx$$

and show that this too can be made arbitrarily small. Now

$$\int\limits_{-a}^{a} |f(x)|dx$$

is a fixed number, say M, and we can take p so large that $\left|\varphi\left(\frac{x}{p}\right) - \varphi(0)\right| < \varepsilon/M$ for $|x/p| \le a$. With this,

$$\left|\int\limits_{-a}^{a} f(x)\left(\varphi\left(\frac{x}{p}\right) - \varphi(0)\right)dx\right| \le \int\limits_{-a}^{a} |f(x)|\left|\varphi\left(\frac{x}{p}\right) - \varphi(0)\right|dx < \varepsilon$$

Combining the three estimates we have

$$\left|\int\limits_{-\infty}^{\infty} f(x)\left(\varphi\left(\frac{x}{p}\right) - \varphi(0)\right)dx\right| < 2\varepsilon$$

and we're done.

We've already seen two applications of this construction, to
$$f(x) = \Pi(x)$$
and, originally, to
$$f(x) = \frac{1}{\sqrt{2\pi}}e^{x^2/2} \quad, take \;\; p = 1/\sqrt{t}.$$
Another possible choice, believe it or not, is
$$f(x) = sinc\; x\,.$$
This works because the integral

$$\int_{-\infty}^{\infty} sinc \; x dx$$

is the Fourier transform of $sinc$ at 0, and you'll recall that we stated the true fact that

$$\int_{-\infty}^{\infty} e^{-2\pi i s t} sinc \; t dt = \begin{cases} 1 & |t| < \dfrac{1}{2} \\ 0 & |t| > \dfrac{1}{2} \end{cases}$$

Now we study The Fourier transform of a tempered distribution. It's time to show how to generalize the Fourier transform to tempered distributions. It will take us oneor two more steps to get to the starting line, but after that it's a downhill race passing effortlessly (almost)through all the important gates.

How to extend an operation from functions to distributions: Try a function first. To definea distribution T is to say what it does to a test function. You give me a test function φ and we have to tellyou$\langle T, \varphi \rangle$— how T operates on φ. We have done this in two cases, one particular and one general. In particular, we defined δ directly by

$$\langle T, \delta \rangle = \varphi(0)$$

In general, we showed how a function f determines a distribution $T_f$by

$$\langle T_f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x) dx$$

provided that the integral exists for every test function. We also say that the distribution comes from afunction. When no confusion can arise we identify the distribution T_f with the function f it comes fromand write

$$\langle f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(\text{x}) dx.$$

When we want to extend an operation from functions to distributions — , when we want to definethe Fourier transform of a distribution, or the reverse of distribution, or the shift of a distribution, or thederivative of a distribution — we take our cue from the way functions determine distributions and askhow the operation works in the case when the pairing is given by integration. What we hope to see is anoutcome that suggests a direct definition (as happened with δ, for example). This is a procedure to follow.It's something to try.

In the following we illustrate the Fourier transform defined .Suppose T is a tempered distribution. Why should such an object have a Fourier transform, and how onearth shall we define it? It can't be an integral, because T isn't a function so there's nothing to integrate.

If $\mathcal{F}T$ is to be itself a tempered distribution (just as $\mathcal{F}\varphi$ is again a Schwartz function if φ is a Schwartz function) then we have to say how $\mathcal{F}T$ pairs with a Schwartz function, because that's what tempered distributions do. So how?

We have a toe-hold here. If ψ is a Schwartz function then $\mathcal{F}\psi$ is again a Schwartz function and we can ask: How does the Schwartz function $\mathcal{F}\psi$ pair with another Schwartz function φ? What is the outcome of $\langle \mathcal{F}\psi, \varphi \rangle$? We know how to pair a distribution that comes from a function ($\mathcal{F}\psi$ in this case) with a Schwartz function;

$$\langle \mathcal{F}\psi, \varphi \rangle = \int_{-\infty}^{\infty} \mathcal{F}\psi(x)\varphi(x)dx.$$

But we can work with the right hand side:

$$\langle \mathcal{F}\psi, \varphi \rangle = \int_{-\infty}^{\infty} \mathcal{F}\psi(x)\varphi(x)dx.$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi ixy} \psi(y)dy \right) \varphi(x)dx$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi ixy} \psi(y) \varphi(x)dydx$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi ixy} \varphi(x)dx \right) \psi(y)dy$$

(the interchange of integrals is justified because $\varphi(x)e^{-2\pi isx}$ and $\psi(x)e^{-2\pi isx}$ are integrable)

$$= \int_{-\infty}^{\infty} \mathcal{F}\varphi(y) \psi(y)dy$$

$$= \langle \psi, \mathcal{F}\varphi \rangle$$

The outcome of pairing $\mathcal{F}\psi$ with φ is:
$\langle \mathcal{F}\psi, \varphi \rangle = \langle \psi, \mathcal{F}\varphi \rangle$.

This tells us how we should make the definition in general:
Let T be a tempered distribution. The Fourier transform of T, denoted by $\mathcal{F}(T)$ or $T\hat{}$, is the tempered distribution defined by
$\langle \mathcal{F}T, \varphi \rangle = \langle T, \mathcal{F}\varphi \rangle$.
for any Schwartz function φ.
This definition makes sense because when φ is a Schwartz function so is $\mathcal{F}\varphi$; it is only then that the pairing $\langle T, \mathcal{F}\varphi \rangle$ is even defined.
We define the inverse Fourier transform by following the same recipe:

Let T be a tempered distribution. The inverse Fourier transform of T, denoted by $\mathcal{F}^{-1}$ (T) orŤ, isdefined by

$$\langle \mathcal{F}^{-1}\mathrm{T}, \varphi \rangle = \langle \mathrm{T}, \mathcal{F}^{-1}\varphi \rangle .$$

for any Schwartz function φ.

Now all of a sudden we have

Fourier inversion:

$$\mathcal{F}^{-1}\mathcal{F}\,\mathrm{T} = \mathrm{T} \ \text{ and } \mathcal{F}\mathcal{F}^{-1}\mathrm{T} = \mathrm{T}$$

for any tempered distribution T.

It's a cinch. Watch. For any Schwartz function φ,

$$\langle \mathcal{F}^{-1}(\mathcal{F}\,\mathrm{T}), \varphi \rangle = \langle \mathcal{F}\,\mathrm{T}, \mathcal{F}^{-1}\varphi \rangle$$
$$= \langle \mathrm{T}, \mathcal{F}\,(\mathcal{F}^{-1}\varphi) \rangle$$

$= \langle \mathrm{T}, \varphi \rangle$(because Fourier inversion works for Schwartz functions)

This says that$\mathcal{F}^{-1}(\mathcal{F}\,\mathrm{T})$ and T have the same value when paired with any Schwartz function. Thereforethey are the same distribution: $\mathcal{F}^{-1}\mathcal{F}\,\mathrm{T} = \mathrm{T}$. The second identity is derived in the same way.

Done. The most important result in the subject, done, in a few lines.

That is, the generalized Fourier transform"contains" the original, classical Fourier transform in the same sense that tempered distributions containclassical functions.

Now we discuss a Fourier transform hit parade. With the definition in place it's time to reap the benefits and find some Fourier transforms explicitly. Wenote one general property

$\mathcal{F}$ is linear on tempered distributions.

This means that

$\mathcal{F}(\mathrm{T}_1 + \mathrm{T}_2) = \mathcal{F}\mathrm{T}_1 + \mathcal{F}\mathrm{T}_2$ and$\mathcal{F}(\alpha \mathrm{T}) = \alpha \mathcal{F}\mathrm{T}$,

α a number. These follow directly from the definition. To wit:

$$\langle \mathcal{F}(\mathrm{T}_1 + \mathrm{T}_2), \varphi \rangle = \langle (\mathrm{T}_1 + \mathrm{T}_2), \mathcal{F}\varphi \rangle = \langle \mathrm{T}_1, \mathcal{F}\varphi \rangle + \langle \mathrm{T}_2, \mathcal{F}\varphi \rangle$$
$$= \langle \mathcal{F}\mathrm{T}_1, \varphi \rangle + \langle \mathcal{F}\mathrm{T}_2, \varphi \rangle = \langle \mathcal{F}\mathrm{T}_1 + \mathcal{F}\mathrm{T}_2, \varphi \rangle$$
$$\langle \mathcal{F}(\alpha\,\mathrm{T}), \varphi \rangle = \langle \alpha\,\mathrm{T}, \mathcal{F}\varphi \rangle = \alpha\langle \mathrm{T}, \mathcal{F}\varphi \rangle = \alpha\langle \mathcal{F}\mathrm{T}, \varphi \rangle = \langle \alpha\mathcal{F}\mathrm{T}, \varphi \rangle$$

The Fourier transform of δ As a first illustration of computing with the generalized Fourier transform we'll find $\mathcal{F}\delta$. The result is:

• The Fourier transform of $\delta$is

$\mathcal{F}\delta = 1$.

This must be understood as an equality between distributions, i.e., as saying that $\mathcal{F}\delta$ and 1 produce thesame values when paired with any Schwartz function φ. Realize that "1" is the constant function, and thisdefines a tempered distribution via integration:

$$\langle 1, \varphi \rangle = \int_{-\infty}^{\infty} 1. \varphi(x) dx$$

That integral converges because $\varphi(x)$ is integrable (it's much more than

integrable, but it's certainly integrable).

We derive the formula by appealing to the definition of the Fourier transform and the definition of δ. On the one hand,

$$\langle \mathcal{F}\delta, \varphi \rangle = \langle \delta, \mathcal{F}\varphi \rangle \mathcal{F} \ \varphi(0) = \int_{-\infty}^{\infty} \varphi(x)dx.$$

On the other hand, as we've just noted,

$$\langle 1, \varphi \rangle = \int_{-\infty}^{\infty} 1.\varphi(x)dx = \int_{-\infty}^{\infty} \varphi(x)dx.$$

The results are the same, and we conclude that $\mathcal{F}\delta = 1$ as distributions. According to the inversion theorem we can also say that $\mathcal{F}^{-1}1 = \delta$.

We can also show that

$$\mathcal{F}1 = \delta.$$

Here's how. By definition,

$\langle \mathcal{F}1, \varphi \rangle = \langle 1, \mathcal{F}\varphi \rangle = \int_{-\infty}^{\infty} \mathcal{F}\varphi(s)ds.$

But we recognize the integral as giving the inverse Fourier transform of $\mathcal{F}\varphi$ at 0:

$$\mathcal{F}^{-1}\mathcal{F}\varphi(t) = \int_{-\infty}^{\infty} e^{2\pi i st}\mathcal{F}\varphi(s)ds \quad and \ at \ t = 0$$

$$\mathcal{F}^{-1}\mathcal{F}\varphi(0) = \int_{-\infty}^{\infty} \mathcal{F}\varphi(s)ds.$$

And now by Fourier inversion on S,

$$\mathcal{F}^{-1}\mathcal{F}\varphi(0) = \varphi(0).$$

Thus

$$\langle \mathcal{F}1, \varphi \rangle = \varphi(0) = \langle \delta, \varphi \rangle$$

and we conclude that $\mathcal{F}1 = \delta$. (We'll also get this by duality and the evenness of δ once we introduce the reverse of a distribution.)

The equations $\mathcal{F}\delta = 1$ and $\mathcal{F}1 = \delta$ are the extreme cases of the trade-off between timelimited and bandlimited signals. δ is the idealization of the most concentrated function possible — it's the ultimate timelimited signal. The function 1, on the other hand, is uniformly spread out over its domain.

It's rather satisfying that the simplest tempered distribution, δ, has the simplest Fourier transform, 1.(Simplest other than the function that is identically zero.) Before there were tempered distributions, however, there was δ, and before there was the Fourier transform of tempered distributions there was $\mathcal{F}\delta = 1$. In the vacuum tube days this had to be established by limiting arguments, accompanied by an uneasiness (among some) over the nature of the limit and what exactly it

produced. Our computation of $\mathcal{F}\delta = 1$ is simple and direct and leaves nothing in question about the meaning of all the quantities involved. Whether it is conceptually simpler than the older approach is something you will have to decide
for yourself.

The Fourier transform of δ_a Recall the distribution δ_a is defined by
$$\langle \delta_a, \varphi \rangle = \varphi(a).$$
What is the Fourier transform of δ_a? One way to obtain $\mathcal{F}\delta_a$ is via a generalization of the shift theorem, even without that we can find $\mathcal{F}\,\delta_a$ directly from the definition, as follows.

The calculation is along the same lines as the one for δ. We have
$$\langle \mathcal{F}\delta_a, \varphi \rangle = \langle \delta_a, \mathcal{F}\varphi \rangle = \mathcal{F}\varphi(a) = \int_{-\infty}^{\infty} e^{-2\pi i a x}\varphi(x)dx.$$

This last integral, which is nothing but the definition of the Fourier transform of φ, can also be interpreted as the pairing of the function $e^{-2\pi i a x}$ with the Schwartz function $\varphi(x)$. That is,
$$\langle \mathcal{F}\delta_a, \varphi \rangle = \langle e^{-2\pi i a x}, \varphi \rangle$$
hence
$$\mathcal{F}\delta_a = e^{-2\pi i s a}.$$
To emphasize once again what all is going on here, $e^{-2\pi i a x}$ is not integrable, but it defines a tempered distribution through
$$\int_{-\infty}^{\infty} e^{-2\pi i a x}\varphi(x)dx$$
which exists because $\varphi(x)$ is integrable. So, again, the equality of $\mathcal{F}\delta_a$ and $e^{-2\pi i s a}$ means they have the same effect when paired with a function in S.

To complete the picture, we can also show that
$$\mathcal{F}e^{2\pi i x a} = \delta_a.$$
(There's the usual notational problem here with variables, writing the variable x on the left hand side. The "variable problem" doesn't go away in this more general setting.) This argument should look familiar: if φ is in S then
$$\langle \mathcal{F}e^{2\pi i a x}, \varphi \rangle = \langle e^{2\pi i a x}, \mathcal{F}\varphi \rangle$$

$$= \int_{-\infty}^{\infty} e^{2\pi i a x}\mathcal{F}\varphi(x)dx$$

(the pairing here is with respect to x)
But this last integral is the inverse Fourier transform of $\mathcal{F}\varphi$ at a, and so we get back $\varphi(a)$. Hence
$$\langle \mathcal{F}e^{2\pi i a x}, \varphi \rangle = \varphi(a) = \langle \delta_a, \varphi \rangle$$
whence

$$\mathcal{F}e^{2\pi iax} = \delta_a.$$

Remark(3.2.5): (Remark on notation)

You might be happier using the more traditional notation $\delta(x)$ for δ and $\delta(x - a)$

for δ_a (and $\delta(x + a)$ for δ_{-a}). we don't have any objection to this — it is a useful notation for manyproblems — but try to remember that the δ -function is not a function and, really, it is not to be evaluated "at points"; the notation $\delta(x)$ or $\delta(x - a)$ doesn't really make sense from the distributional point of view.

In this notation the results so far appear as:

$$\mathcal{F}\delta(x \pm a) = e^{\pm 2\pi isa}, \mathcal{F}e^{\pm 2\pi iax} = \delta(s \mp a)$$

Careful how the $+$ and $-$ enter.

You may also be happier writing

$$\int_{-\infty}^{\infty} \delta(x)\varphi(x)dx = \varphi(0) \quad and \quad \int_{-\infty}^{\infty} \delta(a - x)\varphi(x)dx = \varphi(a).$$

The Fourier transform of sine and cosine We can combine the results above to find the Fouriertransform pairs for the sine and cosine.

$$\mathcal{F}\left(\frac{1}{2}(\delta_a + \delta_{-a})\right) = \frac{1}{2}\left(e^{-2\pi isa} + e^{2\pi isa}\right) = cos2\pi sa.$$

we'll even write the results "at points":

$$\mathcal{F}\left(\frac{1}{2}(\delta(x - a) + \delta(x + a))\right) = cos2\pi sa.$$

Going the other way,

$$\mathcal{F} \, cos \, 2\pi ax = \mathcal{F}\left(\frac{1}{2}\left(e^{2\pi ixa} + e^{-2\pi ixa}\right)\right) = \frac{1}{2}(\delta_a + \delta_{-a}).$$

Also written as

$$\mathcal{F} \, cos \, 2\pi ax = \frac{1}{2}(\delta(s - a) + \delta(s + a))$$

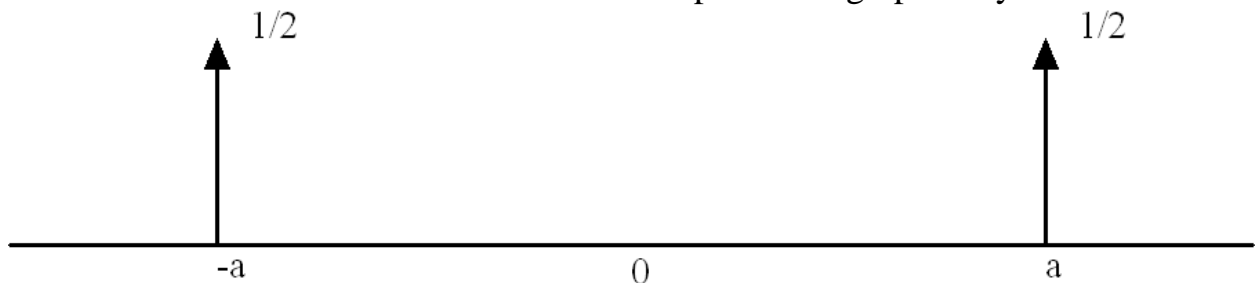The Fourier transform of the cosine is often represented graphically as:

Figure (3.9)

we tagged the spikes with 1/2 to indicate that they have been scaled.

For the sine function we have, in a similar way,

$$\mathcal{F}\left(\frac{1}{2i}(\delta(x + a) - \delta(x - a))\right) = \frac{1}{2i}\left(e^{2\pi isa} - e^{-2\pi isa}\right) = sin2\pi sa \,,$$

121

and

$$\mathcal{F}sin2\pi ax = \mathcal{F}\left(\frac{1}{2i}\left(e^{2\pi ixa} - e^{-2\pi ixa}\right)\right) = \frac{1}{2i}(\delta(s - a) - \delta(s + a))$$
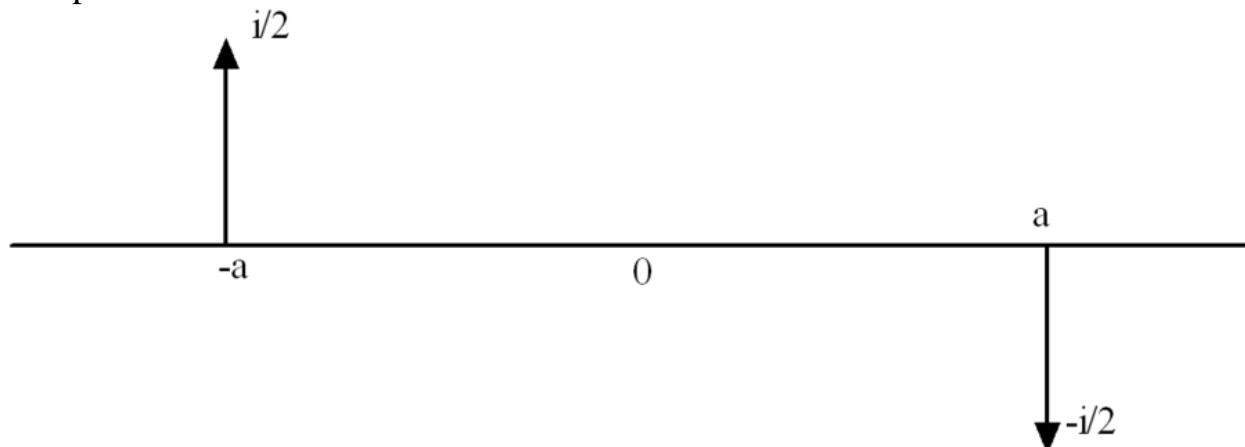
The picture of $\mathcal{F}\ sin2\pi x$ is

Figure (3.10)

Remember that $1/i = -i$. we've tagged the spike δ_a with $-i/2$ and the spike δ_{-a} with $i/2$.

We should reflect a little on what we've done here and not be too quick to move on. The sine and cosinedo not have Fourier transforms in the original, classical sense. It is impossible to do anything with theintegrals

$$\int_{-\infty}^{\infty} e^{-2\pi isx}\ cos2\pi x\ dx \quad or \quad \int_{-\infty}^{\infty} e^{-2\pi isx}\ sin\ 2\pi x\ dx\ .$$

To find the Fourier transform of such basic, important functions we must abandon the familiar, classicalterrain and plant some spikes in new territory. It's worth the effort.

Chapter 4

Distributions and their Fourier Transforms

Section (4.1): Fluxions Finis and Property of the Distributions

We begin this section by study Fluxions Finis: The End of Differential Calculus. For now let's show how introducing distributions "completes" differential calculus; how we can define the derivative of a distribution, and consequently how we can differentiate functions you probably thought had no business being differentiated. We'll make use of this for Fourier transforms, too.

If φ is a test function and f is a function for which $f(x)\,\varphi(x) \to 0$ as $x \to \pm\infty$ (not too much to ask), and if f is differentiable *then* we can use integration by parts to write

$$\int_{-\infty}^{\infty} f'(x)\varphi(x)dx = [f(x)\varphi(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\,\varphi'(x)\left(u = \varphi, dv = f'(x)dx\right)$$

$$= -\int_{-\infty}^{\infty} f(x)\,\varphi'(x)dx.$$

The derivative has shifted from f to φ.

We can find similar formulas for higher derivatives. For example, supposing that the boundary terms in the integration by parts tend to 0 as $x \to \pm\infty$, we find that

$$\int_{-\infty}^{\infty} f''(x)\varphi(x)dx = \left[f'(x)\varphi(x)\right]_{-\infty}^{\infty}$$

$$- \int_{-\infty}^{\infty} f'(x)\,\varphi'(x)\left(u = \varphi(x), dv = f''(x)dx\right)$$

$$= -\int_{-\infty}^{\infty} f'(x)\,\varphi'(x)dx$$

$$= -\left([f(x)\varphi'(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\,\varphi''(x)\right)\left(u = \varphi'(x), dv = f'(x)dx\right)$$

$$= \int_{-\infty}^{\infty} f(x)\,\varphi''(x)dx.$$

Watch out — there's no minus sign out front when we've shifted the *second* derivative from f to φ

We'll concentrate just on the formula for the first derivative. Let's write it again:

$$\int_{-\infty}^{\infty} f'(x)\,\varphi(x)dx = -\int_{-\infty}^{\infty} f(x)\,\varphi'(x)dx\,.$$

The right hand side may make sense even if the left hand side does not, that is, we can view the right hand side as a way of saying how the derivative of f wouldact if it had a derivative. Put in terms of our "try a function first" procedure, if a distribution comes from a function $f(x)$ then this formula tells us how the "derivative" $f'(x)$ *as a* distribution, should be paired with a test function $\varphi(x)$. It should be paired according to the equation above:

$$\langle f',\varphi\rangle = -\langle f,\varphi'\rangle.$$

Turning this outcome into a definition, as our general procedure tells us we should do when passing from functions to distributions, we definethe derivative of a distribution *as* anotherdistribution according to:

If T is a distribution, then its derivative T' is the distribution defined by

$$\langle T',\varphi\rangle = -\langle T,\varphi'\rangle$$

Naturally, $(T_1 + T_2)' = T_1' + T_2'$ $and(\alpha T)' = \alpha T'$. However, there is noproduct rule in general because there's no way to multiply two distributions.

You can go on to define derivatives of higher orders in a similar way, and we'll let you write down what the general formula for the pairing should be. The

striking   thing   is   that   you   don't   have   to   stop:
distributionsareinfinitelydifferentiable.

Derivative of the unit step function The unit step function, also called the
Heaviside functionis defined by

$$H(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

$H(x)$ determines a tempered distribution because for any Schwartz function
$\varphi$the paring

$$\langle H, \varphi \rangle = \int_{-\infty}^{\infty} H(x)\, \varphi(x) dx = \int_{0}^{\infty} \varphi(x) dx$$

makes sense ($\varphi$is integrable).

From the definition of the derivative of a distribution, if $\varphi(x)$ is any test
function then

$$\langle H', \varphi \rangle = -\langle H, \varphi' \rangle = - \int_{-\infty}^{\infty} H(x)\, \varphi'(x) dx = - \int_{0}^{\infty} 1. \varphi'(x) dx$$

$$= -\big(\varphi(\infty) - \varphi(0)\big) = \varphi(0).$$

We see that pairing $H'$with a test function produces the same result as if we had
paired $\delta$with a test function:

$$\langle H', \varphi \rangle = \varphi(0) = \langle \delta, \varphi \rangle.$$

We conclude that

$$H' = \delta.$$

Derivative of the unit ramp, The unit rampfunction is defined by

$$u(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$$

If this were an introductory calculus class and you were asked
"What is the derivative of $u(x)$?" you might have said,
"$It's\ 0\ if\ x \leq 0\ and\ 1\ if\ x > 0$, so it looks like the unit step
$H(x)$ to me." You'd be right, but your jerk of a teacher would

probably say you were wrong because, according to the rigor police, $u(x)$ is not differentiable at $x = 0$. But now that you know about distributions, here's why you were right. For a test function $\varphi(x)$,

$$\langle u'(x), \varphi(x) \rangle = -\langle u(x), \varphi'(x) \rangle = - \int_{-\infty}^{\infty} u(x)\varphi'(x)dx = - \int_{0}^{\infty} x\varphi'(x)dx$$

$$-([x\varphi(x)]_0^{\infty} - \int_{0}^{\infty} \varphi(x)dx) = \int_{0}^{\infty} \varphi(x)dx$$

$(x\varphi(x) \to 0$ as $x \to \infty$ because $\varphi(x)$ decays faster than any power of x)

$$= \langle H, \varphi \rangle$$

Since $\langle u'(x), \varphi(x) \rangle = \langle H, \varphi \rangle$ we conclude that $u' = H$ as distributions.

Then of course, $u'' = \delta$.

Derivative of the signum (or sign) function The signum (or sign) function is

defined by

$$sgn(x) = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \end{cases}$$

Note that sgn is not defined at $x = 0$, but that's not an issue in the derivation to follow.

Let $\varphi(x)$ be any test function. Then

$$\langle sgn', \varphi \rangle = -\langle sgn, \varphi' \rangle = - \int_{-\infty}^{\infty} sgn(x)\varphi'(x)dx$$

$$= -\left(\int_{-\infty}^{0} (-1)\varphi'(x)dx + \int_{0}^{\infty} (+1)\varphi'(x)dx \right)$$

126

$$= \big(\varphi(0) - \varphi(-\infty)\big) - \big(\varphi(\infty) - \varphi(0)\big) = 2\varphi(0)$$

The result of pairing sgn' with φis the same as if we had paired φ $with\ 2\delta$;

$$\langle sgn', \varphi \rangle = 2\varphi(0) = \langle 2\delta, \varphi \rangle$$

Hence

$$sgn' = 2\delta.$$

Observe that $H(x)$ has a unit jump up at 0 and its derivative is δ, whereas sgn jumps up by 2 at 0 and its derivative is 2δ.

Derivative of δTo find the derivative of the δ -function we have, for any test function φ,

$$\langle \delta', \varphi \rangle = -\langle \delta, \varphi' \rangle = -\varphi'(0).$$

That's really as much of a formula as we can write. δ itself acts by pulling out the value of a test function at 0, and δ' acts by pulling out minusthe value of the derivative of the

 test function at 0. we'll let you determine the higher derivatives of δ.

Derivative of$ln|x|$Remember that famous formula from calculus:

$$\frac{d}{dx}\,ln|x| = \frac{1}{x}.$$

Any chance of something like that being true for distributions? Yes, with the properinterpretation. This is an important example because it leads to the *Hilbert transform*, a tool that communications engineers use everyday. For your information, the Hilbert transform is given by convolution of a signal with $1/\pi x$. Once we learn how to take the Fourier transform of $1/x,$ which is coming up, we'll then see that the Hilbert transform is a filter with the interesting property that magnitudes of the spectral components are unchanged but their phases are shifted by $\pm\,\pi/2$.

Because of their usefulness in applications it's worth going through the analysis of the distributions $ln|x|$ and $1/x$.

Now we study Approximations of Distributions and Justifying the "Try a Function First" Principle. We started off by enunciating the principle that to see how to extend an operation from functions to distributions one should start by considering the case when the distribution comes from a function (and hence that the pairing is by integration). Let me offer a justification of why this works.

It's true that not every distribution comes from a function (δ doesn't), but it's also true that any distribution can be approximated by ones that comes from functions. The statement is:

If T is any tempered distribution then there are Schwartz functions f_n such that T_{f_n} converge to T.

This says that for any Schwartz function φ

$$\langle T_{f_n}, \varphi \rangle = \int_{-\infty}^{\infty} f_n\,(x)\varphi(x)dx \to \langle T, \varphi \rangle,$$

that is, the pairing of any tempered distribution with a Schwartz function can be expressed as a limit of the natural pairing with approximating functions via integration. We're not saying that $T_{f_n} \to T_f$ for some function f, because it's not the Schwartz functions f_n that are converging to a function, it's the associated distributions that are converging to a distribution. You don't necessarily have $T = T_f$ for some function f. (Also, this result doesn't say how you're supposed to find the approximating functions, just that they exist.)

Consider how we might apply this to justify our approach to defining the Fourier transform of a tempered distribution. According to the approximation result, any tempered distribution T is a limit of distributions that come from Schwartz functions, and we would have, say,

$$\langle T, \varphi \rangle = \lim_{n \to \infty} \langle \psi_n, \varphi \rangle.$$

Then if $\mathcal{F}T$ is to make sense we might understand it to be given by

$$\langle \mathcal{F}T, \varphi \rangle = \lim_{n \to \infty} \langle \mathcal{F}\psi_n, \varphi \rangle = \lim_{n \to \infty} \langle \psi_n, \mathcal{F}\varphi \rangle = \langle T, \mathcal{F}\varphi \rangle.$$

In the following we discuss The Generalized Fourier Transform Includes the Classical Fourier Transform.Remember that we identify a function f with the distribution T_f it defines and it is in this way we say that the tempered distributions contain many of the classical functions. Now suppose a function $f(x)$ defines a distribution and that $f(x)$ has a (classical) Fourier transform $\mathcal{F}f(s)$ which also defines a distribution, i.e.,

$$\int_{-\infty}^{\infty} \mathcal{F}f(s)\varphi(s)ds$$

exists for every Schwartz function φ(which isn't asking too much). Writing $T_{\mathcal{F}f}$ for the tempered distribution determined by $\mathcal{F}f$,

$$\langle T_{\mathcal{F}f}, \varphi \rangle = \int_{-\infty}^{\infty} \mathcal{F}f(s)\varphi(s)ds$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi isx} f(x)dx \right) \varphi(s)ds = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi isx} f(x)\varphi(s)dsdx$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-2\pi isx} \varphi(s)ds \right) f(x)dx = \int_{-\infty}^{\infty} \mathcal{F}\varphi(x)f(x)dx = \langle T_f, \mathcal{F}\varphi \rangle$$

But now, by our definition of the generalized Fourier transform

$\langle T_f, \mathcal{F}\varphi \rangle = \langle \mathcal{F}T_f, \varphi \rangle$.

Putting this together with the start of the calculation we obtain

$$\langle T_{\mathcal{F}f}, \varphi \rangle = \langle \mathcal{F}T_f, \varphi \rangle,$$

whence

$$T_{\mathcal{F}f} = \mathcal{F}T_f$$

In words, if the classical Fourier transform of a function defines a distribution $(T_{\mathcal{F}f})$ then that distribution is the Fourier transform of the distribution that the function defines $(\mathcal{F}T_f)$.This is a precise way of saying that the generalized Fourier transform "includes" the classical Fourier transform.

Now we discuss Operations on Distributions and Fourier Transforms. We want to relive our past glories — duality between $\mathcal{F}$ and $\mathcal{F}^{-1}$, evenness and oddness, shifts and stretches, convolution — in the more general setting we've developed. The new versions of the old results will ultimately look the same as they did before; it's a question of setting things up properly to apply the new definitions. There will be some new results, however. Among them will be formulas for the Fourier transform of $sgn x$, $1/x$, and the unit step $H(x)$, to take a representative sample.

None of these would have been possible before. We'll also point out special properties of δ along the way. Pay particular attention to these because we'll be using them a lot in applications.

Furthermore, almost all the results are accompanied by some necessary extra notation; the truth is that it's somewhat more cumbersome to define operations on distributions than on functions, and there's no way of getting around it. We have to have this material in some fashion but one should probably treat the sections to follow mostly as a reference. Feel free to use the formulas he need when he need them, and remember that our aim is to recover the formulas we know from earlier work in pretty much the same shape as you first learned them.

In the following we illustrate Duality, Changing Signs, Evenness and Oddness. One of the first things we observed about the Fourier transform and its inverse is that they're pretty much the same thing except for a change in sign. The relationships are

$$\mathcal{F}f(-s) = \mathcal{F}^{-1}f(s)$$
$$\mathcal{F}^{-1}f(-t) = \mathcal{F}f(t)$$

We had similar results when we changed the sign of the variable first and then took the Fourier transform. The relationships are

$$\mathcal{F}(f(-t)) = \mathcal{F}^{-1}f(s)$$
$$\mathcal{F}^{-1}(f(-s)) = \mathcal{F}f(s)$$

We referred to these collectively as the "duality" between Fourier transform pairs, and we'd like to have similar duality formulas when we take the Fourier transforms of distributions.

The problem is that for distributions we don't really have "variables" to change the sign of. We don't really write $\mathcal{F}T(s),$ or $\mathcal{F}T(-s)$ or $T(-s)$, because distributions don't operate on points$s$ — they operate on test functions. What we can do easily is to define a "reversed distribution", and once this is done the rest is plain sailing.

Reversed distributions Recall that we introduced the reversed signal of a signal $f(x)$ by means of

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

and this helped us to write clean, "variable free" versions of the duality results. Using this notation the above results become

$$(\mathcal{F}f)^- = \mathcal{F}^{-1}f, \quad (\mathcal{F}^{-1}f)^- = \mathcal{F}f, \quad \mathcal{F}f^- = \mathcal{F}^{-1}f, \quad \mathcal{F}^{-1}f^- = \mathcal{F}f.$$

A variant version is to apply $\mathcal{F}$ or $\mathcal{F}^{-1}$ twice, resulting in

$$\mathcal{F}\mathcal{F}f = f^-, \qquad \mathcal{F}^{-1}\mathcal{F}^{-1}f = f^-.$$

My personal favorites among formulas of this type are:

$$\mathcal{F}f^- = (\mathcal{F}f)^-, \quad \mathcal{F}^{-1}f^- = (\mathcal{F}^{-1}f)^- .$$

What can "sign change", or "reversal" mean for a distribution T? Our standard approach is first to take the case when the distribution comes from a function $f(x)$. The pairing of T_f with a test function φ is

$$\langle T_f, \varphi \rangle = \int_{-\infty}^{\infty} f(x)\varphi(x)\mathrm{d}x$$

We might well believe that reversing T_f (i.e., a possible definition of $(T_f)^-$ should derive from reversing f, that is, integrating f^- against a test function. The paring of T_{f^-} with φ is

$$\langle T_{f^-}, \varphi \rangle = \int_{-\infty}^{\infty} f(-x)\varphi(x)dx$$

$$= \int_{-\infty}^{\infty} f(u)\varphi(-u)(-du) \; (\textit{making the change of variable } u = -x)$$

$$= \int_{-\infty}^{\infty} f(u)\varphi(-u)du \, .$$

This says that f^- is paired with $\varphi(x)$ in the same way as f is paired with φ^-, more precisely:

$$\langle T_{f^-}, \varphi \rangle = \langle T_f, \varphi^- \rangle$$

Wouldn't it then make sense to say we have found a meaning for $(T_f)^-$ (i.e., have defined $(T_f)^-$) via the formula

$\langle (T_f)^-, \varphi \rangle = \langle T_f, \varphi^- \rangle$ (the right-hand-side is defined because φ^- is defined).

The "outcome" — how this result should be turned into a general definition — is before our eyes:

If T is a distribution we define the reversed distribution T^- according to

$$(T^-, \varphi) = (T, \varphi^-).$$

Note that with this definition we have, quite agreeably,

$$(T_f)^- = T_{f^-}.$$

It's now easy to state the duality relations between the Fourier transform and its inverse. Adopting the notation, above, we want to look at $(\mathcal{F}T)^-$ and how it compares to $(\mathcal{F}^{-1}T)^-$ For a test function φ,

$$((\mathcal{F}T)^-, \varphi) = (\mathcal{F}T, \varphi^-)$$

$= (T, \mathcal{F}(\varphi^-)) \quad (\textit{that's how the Fourier transform is defined})$

$= (T, \mathcal{F}^{-1}\varphi)(\textit{because of duality for ordinary Fourier transforms})$

$$= (\mathcal{F}^{-1}T, \ \varphi)(that's \ how \ the \ inverse \ Fourier \ transform \ is \ defined)$$

Pretty slick, really. We can now write simply

$$(\mathcal{F}T)^- = \ \mathcal{F}^{-1}T.$$

We also then have

$$\mathcal{F}T = (\mathcal{F}^{-1}T)^-.$$

Same formulas as in the classical setting.

To take one more example,

$$\langle \mathcal{F}(T^-), \varphi \rangle = \langle T^-, \mathcal{F}\varphi \rangle = \langle T, (\mathcal{F}\varphi)^- \rangle = \langle T, \mathcal{F}^{-1}\varphi \rangle = \langle \mathcal{F}^{-1}T, \varphi \rangle,$$

and there's the identity

$$\mathcal{F}(T^-) = \mathcal{F}^{-1}T$$

popping out. Finally, we have

$$\mathcal{F}^{-1}(T^-) = \mathcal{F}T.$$

Combining these,

$$\mathcal{F}T^- = (\mathcal{F}T)^- \ , \qquad \mathcal{F}^{-1}T^- = (\mathcal{F}^{-1}T)^-.$$

Applying $\mathcal{F}$ or $\mathcal{F}^{-1}$ twice leads to

$$\mathcal{F}\mathcal{F} \ T = T^-, \quad \mathcal{F}^{-1}\mathcal{F}^{-1}T = T^-.$$

That's all of them.

Even and odd distributions: δ is even Now that we know how to reverse a distribution we can define what it means for a distribution to be even or odd.

·A distribution T is *even* if $T^- = T$. A distribution is *odd* if $T^- = -T$.

Observe that if $f(x)$ determines a distribution T_f and if $f(x)$ is even or odd then T_f has the same property.

 For, as we noted earlier,

$$(T_f)^- = T_{f^-} = T_{\pm f} = \pm T_f.$$

Let's next establish the useful fact:

$\cdot \delta$ is even.

This is quick:

$$\langle \delta', \varphi \rangle = \langle \delta, \varphi' \rangle = \varphi'(0) = \varphi(-0) = \varphi(0) = \langle \delta, \varphi \rangle$$

Let's now use this result plus duality to rederive $\mathcal{F}1 = \delta$. This is quick, too:

$\mathcal{F}1 = (\mathcal{F}^{-1}1) = \delta' = \delta.$

$\delta_a + \delta_{-a}$ is even. $\delta_a - \delta_{-a}$ is odd. Any distribution is the sum of an even and an odd distribution.

You can now show that *all* of our old results on evenness and oddness of a signal and its Fourier transform extend in like form to the Fourier transform of distributions. For example, if T is even then so is $\mathcal{F}T$, for

$(\mathcal{F}T)^- = \mathcal{F}^- = \mathcal{F}T,$
and if T is odd then

$$(\mathcal{F}T)^- = \mathcal{F}T^- = \mathcal{F}(-T) = -\mathcal{F}T,$$

thus $\mathcal{F}T$ is odd.

Notice how this works for the cosine (even) and the sine (odd) and their respective Fourier transforms:

$$\mathcal{F}\cos 2\pi ax = \frac{1}{2}(\delta_a + \delta_{-a})$$

$$\mathcal{F}\sin 2\pi ax = \frac{1}{2i}(\delta_a - \delta_{-a})$$

we'll let you define what it means for a distribution to be real, or purely imaginary.

In the following we study Fourier transform of $\sin c$.

$$\mathcal{F} \sin c = \mathcal{F}(\mathcal{F}\Pi)$$

$$= \Pi^{-}(\text{one of the duality equaltions})$$

$$= \Pi (\Pi \ is \ even)$$

At last. To be really careful here: $\mathcal{F} \sin c$ makes sense only as a tempered distribution. So the equality $\mathcal{F} \sin c = \Pi$ has to be understood as an equation between distributions, meaning that $\mathcal{F} \sin c$ and Π give the same result when paired with any Schwartz function. But you should lose no sleep over this. From now on, write $\mathcal{F} \sin c = \Pi$, think in terms of functions, and start your company.

Now we illustrate a Function Times a Distribution Makes Sense. There's no way todefine the product of two distributions that works consistently with all the rest of the definitions and properties, it just won't work. However, it is possible (and easy) to define the product of a function anda distribution.

Say T is a distribution and g is a function. What is gT as a distribution? we have to tell you what $\langle gT, \varphi \rangle$ is for a test function φ. We take our usual approach to looking for the outcome when T comes from a function, $T = T_f$ .The pairing of gT_f and φ is given by

$$\langle gT_f, \varphi \rangle = \int_{-\infty}^{\infty} g(x)f(x)\,\varphi(x)dx = \int_{-\infty}^{\infty} f(x)(g(x)\,\varphi(x))dx$$

As long as $g\varphi$ is still a test function (so, certainly, g has to be infinitely differentiable) this last integral is the pairing $\langle T_f, g\varphi \rangle$. The outcome is $\langle gT_f, \varphi \rangle = \langle T_f, g\varphi \rangle$. We thus make the following definition:

Definition (4.1.1):

 Let T be a distribution. If g is a smooth function such that $g\varphi$ is a test function whenever φ is a test function, then gT is the distribution defined by

$$\langle gT, \varphi \rangle = \langle T, g\varphi \rangle.$$

This looks as simple as can be, and it is. You may wonder why we even singled out this operation for comment.

In the following we study a function times $\boldsymbol{\delta}$. Watch what happens if we multiply $\boldsymbol{\delta}$ by $\boldsymbol{g(x)}$:

$$\langle g\delta, \varphi \rangle = \langle \delta, g\varphi \rangle = g(0)\varphi(0)$$

This is the same result as if we had paired $g(0)\delta$ with φ. Thus

$$g(x)\delta \ = \ g(0)\delta$$

Inparticularif $g(0) \ = \ 0$ then the result is 0! For example

$$x\delta \ = \ 0$$

or for that matter

$$x^n\delta \ = \ 0$$

for any positive power of x.

Along with $g\delta = g(0)\delta$ we have

$$g(x)\delta_a \ = \ g(a)\delta_a.$$

To show this:

$$\langle g\delta_a, \varphi \rangle = \langle \delta_a, g\varphi \rangle = g(a)\varphi(a) = g(a)\langle \delta_a, \varphi \rangle = \langle g(a)\delta_a, \varphi \rangle.$$

If you want to write this identity more classically, it is

$$g(x)\delta(x - a) \ = \ g(a)\delta(x - a).$$

We'll use this property in many applications, for example when we talk about sampling.

More on a function times δ There's a converse to one of the above properties that's interesting in itself .

If T is a distribution and $xT = 0 \quad then \quad T = c\delta$ for some constant c.

we'll show you the proof of this, but you can skip it if you want. The argument is more involved than the simple statement might suggest, but it's a nice example, and a fairly typical example, of the kind of tricks that are used to prove things in this area. Each to their own tastes.

Knowing where this is going, let me start with an innocent observation. If ψ is a smooth function then

$$\psi(x) = \psi(0) + \int_0^x \psi'(t)dt$$

$= \psi(0) + \int_0^1 x\psi'(xu)du$ (using the substitution $u = t/x$)

$$= \psi(0) + x\int_0^1 \psi'(xu)du \, .$$

Let

$$\Psi(x) = \int_0^1 \psi'(xu)du$$

so that

$$\psi(x) = \psi(0) + x\Psi(x).$$

We'll now use this innocent observation in the case when $\psi(0) = 0$, for then

$$\psi(x) = x\Psi(x).$$

It's clear from the definition of Ψ *that* Ψ is as smooth as ψ is and that if, for example, ψ is rapidly decreasing then so is Ψ. Put informally, we've shown that if $\psi(0) = 0$ we can "factor out an x" and still have a function that's as good as ψ.

Now suppose $xT = 0$, meaning that

$$\langle xT, \varphi \rangle = 0$$

for every test function φ. Fix a smooth windowing function φ_0 that is identically 1 on an interval about $x = 0$, goes down to zero smoothly and is identically zero far enough away from $x = 0$; we mentioned smooth windows earlier .
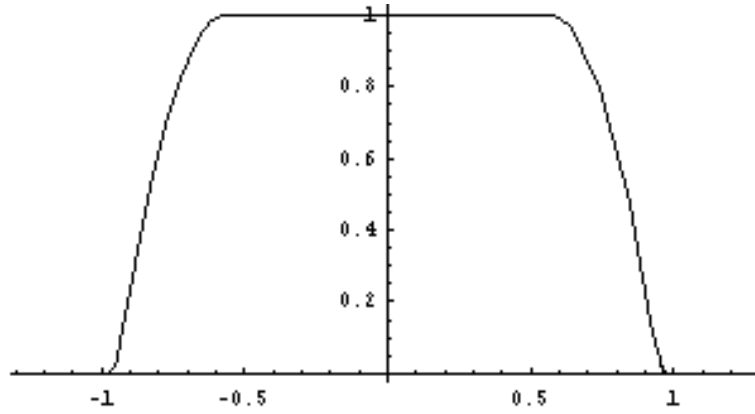
Figure(4.1)

Since φ_0 is fixed in this argument, T operating on φ_0 gives some fixed number, say

$$\langle T, \varphi_0 \rangle = c.$$

Now write

$$\varphi(x) = \varphi(0)\varphi_0(x) + \big(\varphi(x) - \varphi(0)\varphi_0(x)\big)$$
$$= \varphi(0)\varphi_0(x) + \psi(x)$$

where, by this clever way of writing φ , the function $\psi(x) = \varphi(x) - \varphi(0)\varphi_0(x)$ has the property that

$$\psi(0) = \varphi(0) - \varphi(0)\varphi_0(0) = \varphi(0) - \varphi(0) = 0$$

because $\varphi_0(0) = 1$. This means that we can factor out an x and write

$$\psi(x) \ = \ x\Psi(x)$$

where Ψ is again a test function, and then

$$\varphi(x) = \varphi(0)\varphi_0(x) + x\Psi(x).$$

But now

$$\langle T, \varphi(x) \rangle = \langle T, \varphi(0)\varphi_0 + x\Psi \rangle$$
$$= \langle T, \varphi(0)\varphi_0 \rangle + \langle T, x\Psi \rangle$$
$$= \varphi(0)\langle T, \varphi_0 \rangle + \langle T, x\Psi \rangle \text{ (linearity)}$$
$$= \varphi(0)\langle T, \varphi_0 \rangle + \langle xT, \Psi \rangle$$
$$(that's\ how\ mutiplying\ T\ by\ the\ smooth\ function\ x\ works)$$
$$= \varphi(0)\langle T, \varphi_0 \rangle + 0 \qquad (\text{because } \langle xT, \Psi \rangle = 0$$

$$= c\varphi(0)$$
$$= \langle c\delta, \varphi \rangle$$

We conclude that

$$T = c\delta.$$

Now we discuss The Derivative Theorem. Another basic property of the Fourier transform is how it behaves in relation to differentiation — "differentiation becomes multiplication" is the shorthand way of describing the situation. We know how to differentiate a distribution, and it's an easy step to bring the Fourier transform into the picture. We'll then use this to findthe Fourier transform for some common functions that heretofore we have not been able to treat.

Let's recall the formulas for functions, best written:

$$f'(t) \rightleftharpoons 2\pi is \mathrm{F}(s) \quad and \quad -2\pi it f(t) \rightleftharpoons \mathrm{F}'(s)$$

where $f(t) \rightleftharpoons \mathrm{F}(s)$.

We first want to find $\mathcal{F}T'$ for a distribution T. For any test function φ,

$$\langle \mathcal{F}T', \varphi \rangle = \langle T', \mathcal{F}\varphi \rangle = -\langle T, (\mathcal{F}\varphi)' \rangle$$

$$= -\langle T, \mathcal{F}(-2\pi is\varphi) \rangle \text{(from the second formula above)}$$

$$= -\langle \mathcal{F}T, (-2\pi is\varphi) \rangle \text{ (moving } \mathcal{F} \text{ back over to } T)$$

$$= \langle 2\pi is \mathcal{F}T, \varphi \rangle$$

(cancelling minus signs and moving the smooth function $2\pi is$back onto $\mathcal{F}T$)

So the second formula for functions has helped us derive the version of the first

formula for distributions:

$$\mathcal{F}T' = 2\pi is \mathcal{F}T$$

On the right hand side, that's the smooth function $2\pi is$times the distribution $\mathcal{F}T$.

Now let's work with $(\mathcal{F}T)'$:

$$\langle (\mathcal{F}T)', \varphi \rangle = -\langle \mathcal{F}T, \varphi' \rangle = -\langle T, \mathcal{F}(\varphi') \rangle$$

$$= -\langle T, 2\pi i s \mathcal{F}\varphi\rangle \text{(from the first formula for functions)}$$

$$= \langle -2\pi i s T, \mathcal{F}\varphi\rangle$$

$$= \langle \mathcal{F}(-2\pi i s T), \varphi\rangle$$

Therefore

$$(\boldsymbol{\mathcal{F}T})' = \boldsymbol{\mathcal{F}(-2\pi i s T)}.$$

In the following we discuss Fourier transforms of $\boldsymbol{sgn, 1/x}$, and the unit step.We can put the derivative formula to use to find the Fourier transform of the $\boldsymbol{sgn}$ function, and from that the Fourier transform of the unit step.

On the one hand, $sgn' = 2\delta$, from an earlier calculation, so$\mathcal{F}sgn' = 2\mathcal{F}\delta = 2$. On the other hand, using the derivative theorem,

$$\mathcal{F}sgn' = 2\pi i s \mathcal{F}\,sgn\,.$$

Hence

$$2\pi i s \mathcal{F}\,sgn = 2\,.$$

We'd like to say that

$$\mathcal{F}\,sgn = \frac{1}{\pi i s}$$

where $1/s$is the Cauchy principal value distribution. In fact this *is* the case, but it requires a little more of an argument. From $2\pi i s \mathcal{F}\,sgn = 2$ we can say that

$$\mathcal{F}\,sgn = \frac{1}{\pi i s} + c\delta$$

where c is a constant. Why the extra δterm? We need it for generality. If $T$is such that $sT = 0 then\,2\pi i s\,\mathcal{F}\,sgn$ and$2 + sT$, will have the same effect when paired with a test function. But earlier we showed that such a Tmust be $c\delta$for some constant$c$. Thus we write

$$\mathcal{F}\,sgn = \frac{1}{\pi i s} + c\delta$$

Now, sgn is odd and so is its Fourier transform, and so is $\frac{1}{2\pi i s}$. *But* $\delta$is even, and the only way $\frac{1}{\pi i s} + c\delta$can be odd is to have $c = 0$.

To repeat, we have now found

$$\mathcal{F} \text{ sgn} = \frac{1}{\pi i s}.$$

give a derivation of this result using limiting arguments.

By duality we also now know the Fourier transform of $\frac{1}{x}$. The distributions are odd, hence

$$\mathcal{F}\left(\frac{1}{x}\right) = -\pi i \text{ sgn} s.$$

Having found $\mathcal{F}$ sgn it's easy to find the Fourier transform of the unit step H. Indeed,

$$H(t) = \frac{1}{2}(1 + \text{sgn } t)$$

and from this

$$\mathcal{F}H = \frac{1}{2}\left(\delta + \frac{1}{\pi i s}\right).$$

Now we study Shifts and the Shift Theorem. Let's start with shifts. What should we make of $\mathbf{T}(\mathbf{x} \pm \mathbf{b})$ for a distribution $\mathbf{T}$ when, once again, it doesn't make sense to evaluate $\mathbf{T}$ at a point $\mathbf{x} \pm \mathbf{b}$? We use the same strategy as before, starting by assuming that $\mathbf{T}$ comes from a function f and asking how we should pair, say, $f(\mathbf{x} - \mathbf{b})$ with a test function $\boldsymbol{\varphi}(\mathbf{x})$. For that, we want

$$\int_{-\infty}^{\infty} f(x - b)\,\varphi(x)dx = \int_{-\infty}^{\infty} f(u)\boldsymbol{\varphi}(u + b)du$$

(making the substitution $u = x - b$.)

As we did when we analyzed "changing signs" our work on shifts is made easier (really) if we introduce a notation.

The shift or delay operatorIt's pretty common to let τ_b stand for "translate by b", or "delay by b". That is, for any function φ the delayed signal, $\tau_b\varphi$, is the new function defined by

$$(\tau_b\varphi)(x) = \varphi(x - b).$$

Admittedly there's some awkwardness in the notation here; one has to remember that τ_b corresponds to $x - b$.

In terms of τ_b the integrals above can be written (using x as a variable of integration in both cases):

$$\langle \tau_b f, \varphi \rangle = \int_{-\infty}^{\infty} (\tau_b f)(x)\varphi(x)dx = \int_{-\infty}^{\infty} f(x)(\tau_{-b}\varphi(x)dx = \langle f, \tau_{-b}\varphi \rangle.$$

Note that on the left hand side f is shifted by b while on the right hand side φ is shifted by $-b$. This result guides us in making the general definition:

• If $\boldsymbol{T}$ is a distribution we define $\tau_b T$ ($\boldsymbol{T}$ delayed by b) by

$$\langle \tau_b T, \varphi \rangle = \langle T, \tau_{-b}\varphi \rangle.$$

You can check that for a distribution T_f coming from a function f we have

$$\tau_b T_f = T_{\tau_b f} \ .$$

δ_a is a shifted δ To close the loop on some things we said earlier, watch what happens when we delay δ by a:

$$\langle \tau_a \delta, \varphi \rangle = \langle \delta, \tau_{-a}\varphi \rangle$$
$$= (\tau_{-a}\varphi)(0)$$
$$= \varphi(a)(\text{remember}, \tau_{-a}\varphi(x) = \varphi(x+a)$$
$$= \langle \delta_a, \varphi \rangle$$

We have shown that $\tau_a \delta = \delta_a$.

This is the variable-free way of writing $\delta(x - a)$.

Theorem (4.1.2): (The shift theorem)

We're now ready for the general form of the shift theorem:

If $\boldsymbol{T}$ is a distribution then

$$\mathcal{F}(\tau_b T) = e^{-2\pi i b x}\mathcal{F}T \ .$$

To verify this, first

$$\langle \mathcal{F}(\tau_b T), \varphi \rangle = \langle \tau_b T, \mathcal{F}\varphi \rangle = \langle T, \tau_{-b}\mathcal{F}\varphi \rangle.$$

We can evaluate the test function in the last term:

$$\tau_{-b}(\mathcal{F}\varphi)(s) = \mathcal{F}\varphi(s+b)$$

$$= \int_{-\infty}^{\infty} e^{-2\pi i(s+b)x}\,\varphi(x)dx$$

$$= \int_{-\infty}^{\infty} e^{-2\pi isx}e^{-2\pi ibx}\varphi(x)dx$$

$$= \mathcal{F}\big(e^{-2\pi ibx}\varphi\big)(s)$$

Now plug this into what we had before:

$$\langle \mathcal{F}(\tau_b T), \varphi\rangle = \langle T, \tau_{-b}\mathcal{F}\varphi\rangle$$

$$= \langle T, \mathcal{F}\big(e^{-2\pi ibx}\varphi\big)\rangle$$

$$= \langle \mathcal{F}T, e^{-2\pi ibx}\varphi\rangle = \langle e^{-2\pi ibx}\mathcal{F}T, \varphi\rangle$$

Thus, keeping track of what we're trying to show,

$$\langle \mathcal{F}(\tau_b T), \varphi\rangle = \langle e^{-2\pi ibx}\mathcal{F}T, \varphi\rangle$$

for all test functions φ, and hence

$$\mathcal{F}(\tau_b T) = e^{-2\pi ibx}\mathcal{F}T.$$

As one quick application of this let's see what happens to the shifted δ. By the shift theorem

$$\mathcal{F}\tau_a\delta = e^{-2\pi ias}\mathcal{F}\delta = e^{-2\pi isa}$$

in accord with what we found earlier for $\mathcal{F}\delta_a$ directly from the definitions of δ_a and $\mathcal{F}$.

Section (4.2): Stretch Theorem and δ Hard at Work

We begin this section by studying Scaling and the Stretch Theorem. To find the appropriate form of the Stretch Theorem, or Similarity Theorem, we first have to consider how to define $T(ax)$. Following our now usual procedure, we check what happens when T comes from a function f. We need to look at the pairing of $f(ax)$ with a test function

$\varphi(x)$, and we find for $a > 0$ that

$$\int_{-\infty}^{\infty} f(ax)\varphi(x)dx = \int_{-\infty}^{\infty} f(u)\varphi\left(\frac{u}{a}\right)\frac{1}{a}\,du,$$

making the substitution $u = ax$, *and for* $a < 0$ *that*

$$\int_{-\infty}^{\infty} f(ax)\varphi(x)dx = \int_{\infty}^{-\infty} f(u)\varphi\left(\frac{u}{a}\right)\frac{1}{a}\,du = -\int_{-\infty}^{\infty} f(u)\varphi\left(\frac{u}{a}\right)\frac{1}{a}\,du.$$

We combine the cases and write

$$\int_{-\infty}^{\infty} f(ax)\varphi(x)dx = \int_{-\infty}^{\infty} f(u)\varphi\left(\frac{u}{a}\right)\frac{1}{|a|}\,du.$$

The scaling operator, As we did to write shifts in a variable-free way, we do the same for similarities.

We let σ_a stand for the operator "scale by a". That is,

$$\sigma_a(\varphi)(x) = \varphi(ax).$$

The integrals above can then be written as

$$\langle \sigma_a f \varphi \rangle = \int_{-\infty}^{\infty} (\sigma_a f)(x)\varphi(x) = \int_{-\infty}^{\infty} f(x)\frac{1}{|a|}\left(\sigma_{\frac{1}{a}}\varphi\right)(x)dx$$

$$= \langle f, \frac{1}{|a|}(\sigma_{1/a}\varphi) \rangle.$$

144

Thus for a general distribution:

• If T is a distribution we define $\sigma_a T$ via

$$\langle \sigma_a T, \varphi \rangle = \langle T, \frac{1}{|a|} \sigma_{1/a} \varphi \rangle.$$

Note also that then

$$\langle \frac{1}{|a|} \sigma_{1/a} T, \varphi \rangle = \langle T, \sigma_a \varphi \rangle.$$

For a distribution T_f coming from a function f the relation is

$$\sigma_a T_f = T_{\sigma_a f}.$$

Scaling δ *Since* δ is concentrated at a point, however you want to interpret that, you might not think that scaling $\delta(x)$ *to* $\delta(ax)$ should have any effect. But it does:

$$\langle \sigma_a \delta, \varphi \rangle = \langle \delta, \frac{1}{|a|} \sigma_{\frac{1}{a}} \varphi \rangle = \frac{1}{|a|} \left(\sigma_{\frac{1}{a}} \varphi \right)(0)$$

$$= \frac{1}{|a|} \varphi \left(\frac{0}{a} \right) = \frac{1}{|a|} \varphi(0) = \langle \frac{1}{|a|} \delta, \varphi \rangle$$

Hence

$$\sigma_a \delta = \frac{1}{|a|} \delta.$$

This is most often written "at points", as in

$$\delta(ax) = \frac{1}{|a|} \delta(x)$$

The effect of "scaling the variable" is to "scale the strength" of δ by the reciprocal amount.

The stretch theorem With the groundwork we've done it's now not difficult to state and derive the general stretch theorem:

If T is a distribution then

$$\mathcal{F}(\sigma_a T) = \frac{1}{|a|}\sigma_{\frac{1}{a}}(\mathcal{F}T).$$

To check this,

$$\langle \mathcal{F}(\sigma_a T), \varphi \rangle = \langle \sigma_a T, \mathcal{F}\varphi \rangle = \langle T, \frac{1}{|a|}\sigma_{\frac{1}{a}}\mathcal{F}\varphi \rangle.$$

But now by the stretch theorem for functions

$$\frac{1}{|a|}\left(\sigma_{\frac{1}{a}}\mathcal{F}\varphi\right)(s) = \frac{1}{|a|}\mathcal{F}\varphi(\frac{s}{a}) = \mathcal{F}(\sigma_a \varphi)(s).$$

Plug this back into what we had:

$$\langle \mathcal{F}(\sigma_a T), \varphi \rangle = \langle T, \frac{1}{|a|}\sigma_{\frac{1}{a}}\mathcal{F}\varphi \rangle$$

$$\langle T, \mathcal{F}(\sigma_a \varphi) \rangle = \langle \mathcal{F}T, \sigma_a \varphi \rangle = \langle \frac{1}{|a|}\sigma_{\frac{1}{a}}(\mathcal{F}T), \varphi \rangle.$$

This proves that

$$\mathcal{F}(\sigma_a T) = \frac{1}{|a|}\sigma_{\frac{1}{a}}(\mathcal{F}T).$$

Now we study Convolutions and the Convolution Theorem. Convolution of distributions presents some special problems and we're not going to go into this too deeply. It's not so hard figuring out formally how to define $S * T$ for distributions S and T, it's setting up conditions under which the convolution exists that's somewhat tricky. This is related to the fact of nature that it's impossible to define (in general) the product of two distributions, for we also want to have a convolution theorem that says

$\mathcal{F}(S * T) = (\mathcal{F}S)(\mathcal{F}T)$ and both sides of the formula should make sense.

What works easily is the convolution of a distribution with a test function. This goes through as you might expect (with a little twist) but in case you want to skip the following discussion I am pleased to report right away that the convolution theorem on Fourier transforms continues to hold: If ψ is a test function and T is a distribution then

$$\mathcal{F}(\psi * T) = (\mathcal{F}\psi)(\mathcal{F}T).$$

The right hand side is the product of a test function and a distribution, which isdefined.

Here's the discussion that supports the development of convolution in this setting. First we consider how to define convolution of ψ and T. As in every other case of extending operations from functions to distributions, we suppose first that a distribution T comes from a function f. If ψ is a test function we want to look at the pairing of $\psi * f$ with a test function φ. This is

$$
\langle \psi * f, \varphi \rangle = \int_{-\infty}^{\infty} (\psi * f)(x)\varphi(x)dx
$$

$$
= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \psi(x - y)f(y)\,dy \right)\varphi(x)dx
$$

$$
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x - y)\varphi(x)f(y)\,dydx
$$

$$
= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \psi(x - y)\,\varphi(x)dx \right)f(y)dy
$$

(The interchange of integration in the last line is justified because every function in sight is as nice as can be.) We almostsee a convolution $\psi * \varphi$ in the inner integral — but the sign is wrong. However, bringing back our notation $\psi^-(x) = \psi(-x)$, we canwrite the inner integral as the convolution $\psi^- * \varphi$ (or as $\psi * \varphi^-$ by a change of variable). That is

$$
\langle \psi * f, \varphi \rangle = \int_{-\infty}^{\infty} (\psi * f)(x)\varphi(x)dx = \int_{-\infty}^{\infty} (\psi^- * \varphi)(x)f(x)dx = \langle f, \psi^- * \varphi \rangle.
$$

This tells us what to do in general:

• If T is a distribution and ψ is a test function then $\psi * T$ is defined by

$$
\langle \psi * T, \varphi \rangle = \langle T, \psi^- * \varphi \rangle.
$$

Convolution property ofδLet's see how this works to establish the basic convolution property of the δ-function:

$$
\psi * \delta = \psi
$$

where on the right hand side we regard ψ as a distribution. To check this:

$$\langle \psi * \delta, \varphi \rangle = \langle \delta, \psi^- * \varphi \rangle = (\psi^- * \varphi)(0)$$

Look at this carefully, or rather, simply. It says that $\psi * \delta$ has the same outcome as ψ does when paired with ϕ. That is, $\psi * \delta = \psi$. Works like a charm. Air tight.

As pointed out earlier, it's common practice to write this property of δ as an integral,

$$\psi(x) = \int_{-\infty}^{\infty} \delta(x-y)\psi(y)dy.$$

This is sometimes called the sifting property of δ. Generations of distinguished engineers and scientists have written this identity in this way, and no harm seems to have befallen them.

We can even think of Fourier inversion as a kind of convolution identity, in fact as exactly the sifting property of δ. The inversion theorem is sometimes presented in this way (proved, according to some people, though it's circular reasoning). We need to write (formally)

$$\int_{-\infty}^{\infty} e^{2\pi i s x}\, ds = \delta(x)$$

viewing the left hand side as the inverse Fourier transform of 1, and then, shifting,

$$\int_{-\infty}^{\infty} e^{2\pi i s x} e^{-2\pi i s t}\, ds = \delta(x-t)$$

And now, shamelessly,

$$\mathcal{F}^{-1}\mathcal{F}\varphi(x) = \int_{-\infty}^{\infty} e^{2\pi i s x}\left(\int_{-\infty}^{\infty} e^{-2\pi i s t}\, \varphi(t)dt\right) ds$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2\pi i s x} e^{-2\pi i s t}\, \varphi(t)dt\, ds$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{2\pi i s x} e^{-2\pi i s t}\, ds\right)\varphi(t)dt$$

$$= \int_{-\infty}^{\infty} \delta(x - t)\varphi(t)dt = \varphi(x).$$

Having come this far, we can now derive the convolution theorem for the Fourier transform:

$$\langle \mathcal{F}(\psi * T), \varphi \rangle = \langle \psi * T, \mathcal{F}\varphi \rangle = \langle T, \psi^- * \mathcal{F}\varphi \rangle$$

$$= \langle T, \mathcal{F}\mathcal{F}\psi * \mathcal{F}\varphi \rangle \ (using \ the \ identity \ \mathcal{F}\mathcal{F}\psi = \psi^-)$$

$$= \langle T, \mathcal{F}(\mathcal{F}\psi.\varphi) \rangle$$

$$(for \ functions \ the \ convolution \ of \ the \ Fourier \ transfoms \ is \ the$$
$$Fourier \ transform \ of \ the \ product)$$
$$= \langle \mathcal{F}T, \mathcal{F}\psi.\varphi \rangle \ (bringing \ \mathcal{F} \ back \ to \ T)$$
$$= \langle (\mathcal{F}\psi)(\mathcal{F}T), \varphi \rangle (how \ multiplication \ by \ a \ function \ is \ defined)$$

Comparing where we started and where we ended up:

$$\langle \mathcal{F}(\psi * T), \varphi \rangle = \langle (\mathcal{F}\psi)(\mathcal{F}T), \varphi \rangle.$$

that is,

$$\mathcal{F}(\psi * T) = (\mathcal{F}\psi)(\mathcal{F}T).$$

Done.

One can also show the dual identity:

$$\mathcal{F}(\psi T) = \mathcal{F}\psi * \mathcal{F}T$$

Pay attention to how everything makes sense here and has been previously defined. The product of the Schwartz function ψ and the distribution T is defined, and as a tempered distribution it has a Fourier transform. Since ψ is a Schwartz function so is its Fourier transform $\mathcal{F}\psi$, and hence $\mathcal{F}\psi * \mathcal{F}T$ is defined.

we'll leave it to you to check that the algebraic properties of the convolutioncontinue to hold for distributions, whenever all the quantities are defined.

Note that the convolution identities are consistent with$\psi * \delta = \psi$, and with

$\psi\delta = \psi(0)\delta$. The first of these convolution identities says that

$$\mathcal{F}(\psi * \delta) = \mathcal{F}\psi\mathcal{F}\delta = \mathcal{F}\psi,$$

since $\mathcal{F}\delta = 1$, and that jibes with $\psi * \delta = \psi$. The other identity is a little more interesting. We have

$$\mathcal{F}(\psi\delta) = \mathcal{F}\psi * \mathcal{F}\delta = \mathcal{F}\psi * 1 = \int_{-\infty}^{\infty} 1.\mathcal{F}(\psi)dx = \mathcal{F}^{-1}\mathcal{F}\psi(0) = \psi(0)$$

This is consistent with $\mathcal{F}(\psi\delta) = \mathcal{F}(\psi(0)\delta) = \psi(0)\mathcal{F}\delta = \psi(0)$.

Convolution in general:we said earlier that convolution can't be defined for every pair of distributions. We want to say a little more about this, but only a little, and give a few examples of cases when it works out.

At the beginning of this we considered, as we always do, what convolutionlooks like for distributions in the case when the distribution comes from a function. With f playing the role of the distribution andψa Schwartz function we wrote

$$\langle \psi * f, \varphi \rangle = \int_{-\infty}^{\infty} (\psi * f)(x)\varphi(x)dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \psi(x-y)f(y)dy \right)\varphi(x)dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x-y)\varphi(x)f(y)dxdy$$

$$= \int_{-\infty}^{\infty} (\int_{-\infty}^{\infty} \psi(x-y)\varphi(x)dx)\, f(y)dy.$$

At this point we stopped and wrote this as the pairing

$$\langle \psi * f, \varphi \rangle = \langle f, \psi^- * \varphi \rangle$$

so that we could see how to define $\psi * T$when T is a distribution.

This time, and for a different reason, we want to take the inner integral one step further and write

$$\int_{-\infty}^{\infty} \psi(x-y)\varphi(x)dx = \int_{-\infty}^{\infty} \psi(u)\varphi(u+y)du$$

$$(using\ the\ substituion\ u\ =\ x\ -\ y).$$

This latter integral is the pairing $\langle \psi(x), \varphi(x + y) \rangle$, where we wrote the variable of the paring (the integration variable) as x and I included it in the notation for pairing to indicate that what results from the pairing is a function y. In fact, what we see from this is that $\langle \psi * f, \varphi \rangle$ can be written as a "nested" pairing, namely

$$\langle \psi * f, \varphi \rangle = \langle f(y), \langle \psi(x), \varphi(x + y) \rangle \rangle$$

where we included the variable y in the outside pairing to keep things straight and to help recall that in the end everything gets integrated away and the result of the nested pairing is a number.

Now, this nested pairing tells us how we might define the convolution $S * T$ of two distributions $S\ and\ T$. It is, with a strong proviso:

Convolution of two distributions If S and T are two distributions then their convolution is the distribution $S * T$ defined by

$$\langle S * T, \varphi \rangle = \langle S(y), \langle T(x), \varphi(x + y) \rangle \rangle$$

provided the right-hand-side exists.

We've written $S(y)\ and\ T(x)$ "at points" to keep straight what gets paired with what; $\varphi(x + y)$ makes sense, is a function of $x\ and\ y,$ and it's necessary to indicate which variable $x\ or\ y$ is getting hooked up with T in the inner pairing and then with S in the outer pairing.

Why the proviso? Because the inner paring $\langle T(x), \varphi(x + y) \rangle$ produces a function of y which might not be a test function. Sad, but true. One can state some general conditions under which $S * T$ exists, but this requires a few more definitions and a little more discussion. Enough is enough. It can be dicey, but we'll play a little fast and loose with existence of convolution and applications of the convolution theorem. Tell the rigor police to take the day off.

Convolving δ with itself. For various applications you may find yourself wanting to use the identity

$$\delta * \delta = \delta.$$

By all means, use it. In this case the convolution makes sense and the formula follows:

$$\langle \delta * \delta, \varphi \rangle = \langle \delta(y), \langle \delta(x), \varphi(x+y) \rangle \rangle$$

$$\langle \delta(y), \varphi(y) \rangle = \varphi(0) = \langle \delta, \varphi \rangle.$$

A little more generally, we have

$$\delta_a * \delta_b = \delta_{a+b},$$

a nice formula! We can derive this easily from the definition:

$$\langle \delta_a * \delta_b, \varphi \rangle = \langle \delta_a(y), \langle \delta_b(x), \varphi(x+y) \rangle \rangle$$

$$= \langle \delta_a(y), \varphi(b+y) \rangle = \varphi(b+a) = \langle \delta_{a+b}, \varphi \rangle.$$

It would be more common to write this identity as

$$\delta(x-a) * \delta(x-b) = \delta(x-a-b).$$

In this notation, here's the down and dirty version of what we just did (so you know how it looks):

$$\delta(x-a) * \delta(x-b) = \int_{-\infty}^{\infty} \delta(y-a)\, \delta(x-b-y)dy$$

$$= \int_{-\infty}^{\infty} \delta(u-b-a)\, \delta(x-u)du \quad (using\ u = b+y)$$

$$= \delta(x-b-a)(by\ the\ sifting\ property\ of\ \delta).$$

Convolution really is a "smoothing operation" (most of the time).we want to say a little more about general properties of convolution (first for functions) and why convolution is a smoothing operation. In fact, it's often taken as a maxim when working with convolutions that:

• The function $f * g$ has the good properties of $f\ and\ g$.

This maxim is put to use through a result called the derivativetheoremforconvolutions:

$$(f * g)'(x) = (f * g')(x) = (f' * g)(x).$$

On the left hand side is the derivative of the convolution, while on the right hand side we put the derivative on whichever factor hasa derivative.

We allow ourselves to differentiate under the integral sign — sometimes a delicate business, but set that aside — and the derivation is easy. If g is differentiable, then

$$(f * g)'(x) = \frac{d}{dx} \int_{-\infty}^{\infty} f(u)g(x-u)\,du$$

$$= \int_{-\infty}^{\infty} f(u) \frac{d}{dx} g(x-u)\,du = \int_{-\infty}^{\infty} f(u)g'(x-u)\,du = (f * g')(x)$$

The second formula follows similarly if f is differentiable.

The importance of this is that the convolution of two functions may have more smoothness than the individual factors. We've seen one example of this already, where it's not smoothness but continuity that's improved. Remember $\Pi * \Pi = \Lambda$; the convolution of the rectangle function with itself is the triangle function. The rectangle function is not continuous — it has jump discontinuities at $x = \pm 1/2$ — but the convolved function is continuous. We also saw that repeated convolution of a function with itself will lead to a Gaussian.

The derivative theorem is saying: If f is rough, but g is smooth then $f * g$ will be smoother than f because we can differentiate the convolution by putting the derivative on g. We can also compute higher order derivatives in the same way. If g is n-times differentiable then

$$(f * g)^n(x) = (f * g^n)(x).$$

Thus convolving a rough function f with an n-times differentiable function g produces an n-times differentiable function $f * g$. It is in this sense that convolution is a "smoothing" operation.

The technique of smoothing by convolution can also be applied to distributions. There one works with $\psi * T$ where ψ is, for example, a Schwartz function. Using the family of Gaussians $g_t = \left(\frac{1}{\sqrt{2\pi t}}\right) e^{-\frac{x^2}{2t}}$ to form $g_t * T$ produces the so-called regularization of T. This is the basis of the theorem on approximating a general distribution by a sequence of distributions that come from Schwartz functions. The distribution δ is the breakeven point for smoothing by convolution — it doesn't do any smoothing, it leaves the function alone, as in

$$\delta * f = f.$$

Going further, convolving a differentiable function with derivatives of δ produces derivatives of the function, for example,

$$\delta' * f = f'.$$

we can derive this from scratch using the definition of the derivative of a distribution and the definition of convolution, or we can also think of

$$\delta' * f = \delta * f' = f'.$$

(Careful here: This is δ' convolved with f, not δ' paired with f.) A similar result holds for higher derivatives:

$$\delta^{(n)} * f = f^{(n)}.$$

Sometimes one thinks of taking a derivative as making a function less smooth, so counterbalancing the maxim that convolution is a smoothing operation, one should add that convolving with derivatives of δ may roughen a function up.

in the following we discuss δ Hard at work We've put a lot of effort into general theory and now it's time to see a few applications. They range from finishing some work on filters, to optics and diffraction, to X-ray crystallography. The latter will even lead us toward the sampling theorem. The one thing all these examples have in common is their use of $\delta's$.

The main properties of δ we'll need, along with its Fourier transform, are what happens with convolution with a function ϕ and with multiplication by a function φ:

$$\delta * \varphi = \varphi \quad and \quad \varphi \delta = \varphi(0)\delta.$$

We'll tend to "write the variables" in this section, so these identities appear as

$$\int_{-\infty}^{\infty} \delta(x - y)\varphi(y)dy = \varphi(x) \, and \quad \varphi(x)\delta(x) = \varphi(0)\delta(x).$$

There are useful variations of these formulas for a shifted δ:

$$\delta(x - b) * \varphi(x) = \varphi(x - b)$$

$$\delta(x - b)\varphi(x) = \varphi(b)\delta(x - b)$$

We also need to recall the Fourier transform for a scaled rect:

$$\mathcal{F}\Pi_a(x) = \mathcal{F}\Pi(x/a) = a \, sin \, ca.$$

Now we illustrate Filters, redux. One of our first applications of convolution was to set up and study some simple filters. Let's recall the terminology and some work left undone; The input $v(t)$ and the output $w(t)$ are related via convolution with the impulse response $h(t)$:

$$w(t) = (h * v)(t).$$

(We're not quite ready to explain why h is called the impulse response.) The action of the filter is easier to understand in the frequency domain, for there, by the convolution theorem, it acts by multiplication

$$W(s) = H(s)V(s)$$

where

$$W = \mathcal{F}w, H = \mathcal{F}h, and \ V = \mathcal{F}v.$$

$H(s)$ is called the transferfunction.

The simplest example, out of which the others can be built, is the low-pass filter with transfer function

$$Low(s) = \Pi_{2v_c}(s) = \Pi\left(\frac{s}{2v_c}\right) = \begin{cases} 1, & |s| < v_c \\ 0, & |s| \geq v_c \end{cases}$$

The impulse response is

$$Low(t) = 2v_c sin(2v_c t)$$

a scaled $sin c$ function.

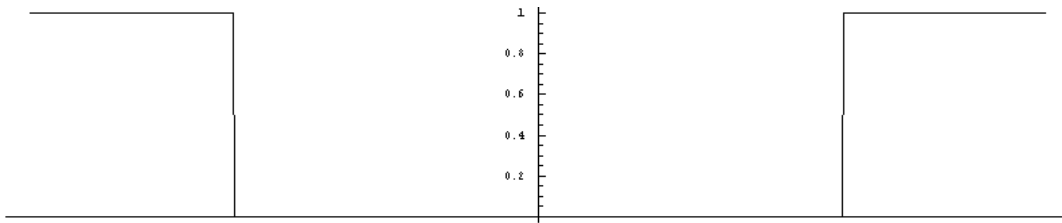High-pass filter, earlier we saw the graph of the transfer function for an ideal high pass filter:

Figure (4.2)

and a formula for the transfer function

$$High(s) = 1 - Low(s) = 1 - \Pi_{2v_c}(s)$$

where v_c is the cut-off frequency. At the time we couldn't finish the analysis because we didn't have δ. Now we do. The impulse response is

$$high(t) = \delta(t) - 2v_c sinc(2v_c t).$$

For an input $v(t)$ the output is then

$$w(t) = (high * v)(t)$$

$$= \left(\delta(t) - 2v_c sinc(2v_c t)\right) * v(t)$$

$$= v(t) - 2v_c \int_{-\infty}^{\infty} sinc\left(2v_c(t - s)\right)v(s)ds.$$

The role of the convolution property of δ in this formula shows us that the high pass filter literally subtracts part of the signal away.

Notch filter The transfer function for the notch filter is just $1-$ (transfer function for band pass filter) and it looks like this:
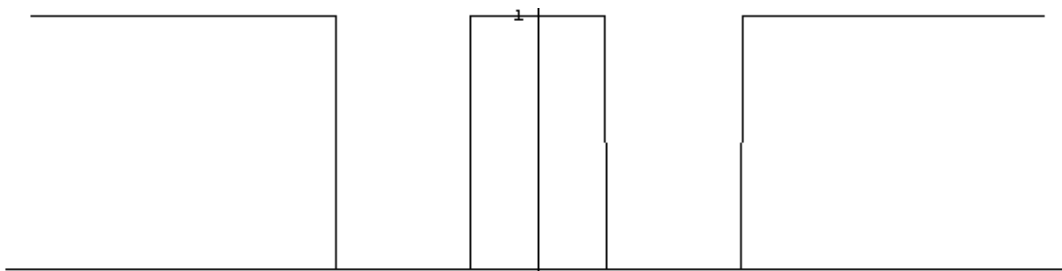

Figure (4.3)

Frequencies in the "notch" are filtered out and all others are passed through unchanged. Suppose that the notches are centered at $\pm v_0$ and that they are v_c wide. The formula for the transfer function, in terms of transfer function for the low-pass filter with cutoff frequency v_c, is

$$Notch(s) = 1 - \big(Low(s - v_0) + Low(s + v_0)\big).$$

For the impulse response we obtain

$$notch(t) = \delta(t) - (e^{-2\pi i v_0 t} low(t) + e^{2\pi i v_0 t} low(t))$$

$$= \delta(t) - 4v_c cos(2\pi v_0 t) sinc(2v_c t).$$

Thus

$$w(t) = (\delta(t) - 4v_c cos(2\pi v_0 t) sinc(2v_c t)) * v(t)$$

$$= v(t) - 4v_c \int_{-\infty}^{\infty} cos(2\pi v_0 (t - s)) sinc(2v_c (t - s)) v(s) ds,$$

and again we see the notch filter subtracting away part of the signal.

In the following we study Diffraction: The **sinc** function, live and in pure color. Some of the most interesting applications of the Fourier transform are in the field of optics, understood broadly to include most of the electromagnetic spectrum in its purview.

The fundamental phenomenon associated with the wave theory of light is diffractionor interference. Sommerfeld says that diffraction is "any deviation of light rays from rectilinear paths which cannot be interpreted as reflection or refraction." Very helpful. Is there a difference between diffraction and interference?. It is just a question of usage, and there is no specific, important physical difference between them." He does go on to say that "interference" is usually associated with patterns caused by a few radiating sources, like two, while "diffraction" is due to many sources. Whatever the definition, or nondefinition, you probably know what the picture is:
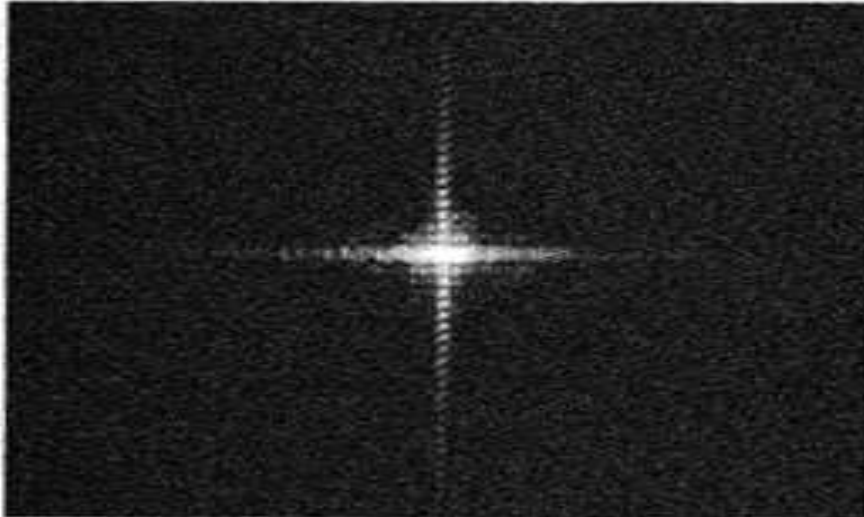
Figure (4.4)

Such pictures, most notably the "Two Slits" experiments of Thomas Young (1773–1829), which we'll analyze, below, were crucial in tipping the balance away from Newton's corpuscular theory to the wave theory propounded by Christiaan Huygens (1629–1695). The shock of the diffraction patterns when first seen was that light + light could be dark. Yet the experiments were easy to perform. Spoke Young in 1803 to the Royal Society: "The experiments we about to relate ... may be repeated with great ease, whenever the sun shines, and without any other apparatus than is at hand to every one."

We are thus taking sides in the grand battle between the armies of "light is a wave" and those of "light is a particle". It may be that light is truly like nothing you've ever seen before, but for this discussion it's a wave. Moreover, jumping ahead to Maxwell, we assume that light is an electromagnetic wave, and for our discussion we assume further that the light in our problems is:

1.   Monochromatic

◦ Meaning that the periodicity in time is a single frequency, so described by a simple sinusoid.

2.   Linearly polarized

◦ Meaning that the electric field vector stays in a plane as the wave moves. (Hence so too does the magnetic field vector.)With this, the diffraction problem can be stated as follows:

Light — an electromagnetic wave — is incident on an (opaque) screen with one or more apertures (transparent openings) of various shapes. What is the intensity of the light on a screen some distance from the diffracting screen?

We're going to consider only a case where the analysis is fairly straightforward, the Fraunhoferapproximation, or Fraunhofer diffraction. This involves a number of simplifying assumptions, but the results are used widely. Before we embark on the analysis let me point out that reasoning very similar to what we'll do here is used to understand the radiation patterns of antennas.

We discuss Light waves.We can describe the properties of lightthat satisfythe above assumptions by a scalar-valued function of time and position. We're going to discuss "scalar" diffraction theory, while more sophisticated treatments handle the "vector" theory. The function is the magnitudeof the electric field vector, say a function of the form

$$u(x, y, z, t) = a(x, y, z)cos(2\pi vt - \phi(x, y, z))$$

Here, $a(x, y, z)$is the amplitude as a function only of position in space, $v$is the (single) frequency, and $\phi(x, y, z)$ is the phase at $t = 0$, also as a function only of position.

The equation

$$\phi(x, y, z) = constant$$

describes a surface in space. At a fixed time, all the points on such a surface have the same phase, by definition, or we might say equivalently that the traveling wave reaches all points of such a surface $\varphi(x, y, z) = $ constant at the same time. Thus any one of the surfaces $\phi(x, y, z) = constant$ is called a wavefront. In general, the wave propagates through space in a direction normal to the wavefronts.

The function $u(x, y, z, t)$ satisfies the 3-dimensional wave equation

$$\Delta u = \frac{1}{c^2}\frac{d^2u}{d\,t^2}$$

where

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is the Laplacian and c is the speed of light in vacuum. For many problems it's helpful to separate the spatial behavior of the wave from its temporal behavior and to introduce the complexamplitude, defined to be

$$u(x,y,z) \;=\; a(x,y,z)e^{i\phi(x,y,z)}.$$

Then we get the time-dependent function $u(x,y,z,t)$ as

$$u(x,y,z,t) \;=\; Re\left(\overline{u(x,y,z)}e^{2\pi i v t}\right).$$

If we know $u(x,y,z)$ we can get $u(x,y,z,t)$ It turns out that $u(x,y,z)$ satisfies the differential equation

$$\Delta u(x,y,z) + k^2 u(x,y,z) \;=\; 0$$

where $k = 2\pi v/c$. This is called the Helmholtz equation, and the fact that it is time independent makes it simpler than the wave equation.

Fraunhofer diffractionWe take a sideways view of the situation. Light is coming from a source at a point O and hits a plane S. We assume that the source is so far away from S that the magnitude of the electric field associated with the light is constant on S and has constant phase, i.e., S is a wavefront and we have what is called a planewavefield. Let's say the frequency is v and the wavelength is λ. Recall that $c = \lambda v$, where c is the speed of light. (We're also supposing that the medium the light is passing through is isotropic, meaning that the light is traveling at velocity c in any direction, so there are no special effects from going through different flavors of jello or something like that.)

Set up coordinates so that the $z - axis$ is perpendicular to S and the $x - axis$ lies in S, perpendicular to the $z - axis$. (In most diagrams it is traditional to have the $z - axis$ be horizontal and the $z - axis$ be vertical.)

In S we have one or more rectangular apertures. We allow the length of the side of the aperture along the $x - axis$ to vary, but we assume that the other side (perpendicular to the plane of the diagram) has length 1.

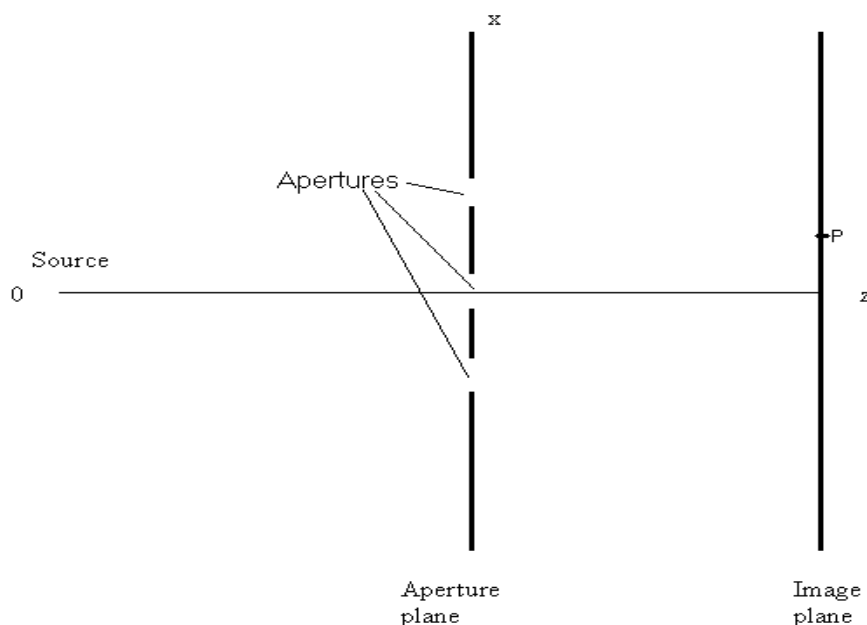 A large distance from S is another parallel plane. Call this the image plane.

Figure (4.5)

The diffraction problem is:

• What is the electric field at a point P in the image plane?. The derivation we going to give to answer this question is not as detailed as is possible but we'll get the correct form of the answer and the point is to see how the Fourier transform enters.

The basis for analyzing diffraction is Huygens'principlewhich states, roughly, that the apertures on S(which is a wavefront of the original source) may be regarded as (secondary) sources, and the field at P is the sum (integral) of the fields coming from these sources on S. Putting in a little more symbolism, if $E_0$is the strength of the electric field on Sthen an aperture of area $dS$is a source of strength $dE = E_0 dS$.

At a distance $r$from this aperture the field strength is $dE^{''} = E_0 dS/r$, and we get the electric field at this distance by integrating over the apertures the elements $dE^{''}$, "each with its proper phase". Let's look more carefully at the phase.

The wave leaves a point on an aperture in S, a new source, and arrives at p sometime later. Waves from different points on S will arrive at p at different times, and hence there will be a phase difference between the arriving waves. They also drop off in amplitude like one over the distance to p, and so by different amounts, but if, as we'll later assume, the size of the apertures on s are small compared to the distance between S and the image plane then this is not as significant as the phase differences. Light is moving so fast that even a small differences between locations of secondary point sources on s may lead to significant differences in the phases when the waves reach p.

The phase on S is constant and we might as well assume that it's zero. Then we write the electric field on S in complex form as

$$E = E_0 e^{2\pi i v t}$$

where E_0 is constant and v is the frequency of the light. Suppose p is at a distance r from a point x on s. Then the phase change from x to p depends on how big r is compared to the wavelength λ— how many wavelengths (or fractions of a wavelength) the wave goes through in going a distance r from x to p.

This is $2\pi(r/\lambda)$. To see this, the wave travels a distance r in a time r/c seconds, and in that time it goes through $v(r/c)$ cycles. Using $c = \lambda v$ that's $vr/c = r/\lambda$. This is $2\pi r/\lambda$ radians, and that's the phase shift.

Take a thin slice of width dx at a height x above the origin of an aperture on s. Then the field at p due to this source is, on account of the phase change,

$$dE = E_0 e^{2\pi i v t} e^{2\pi i r/\lambda} dx.$$

The total field at p is
$$\int_{apertures} E_0 e^{2\pi i v t} e^{2\pi i r/\lambda} dx. = E_0 e^{2\pi i v t} \int_{apertures} e^{2\pi i r/\lambda} dx$$
There's a Fourier transform coming, but we're not there yet.

The key assumption that is now made in this argument is to suppose that

$$r \gg x,$$

that is, the distance between the plane s and the image plane is much greater than any x in any aperture, in particular r is large compared to any aperture size. This assumption is what makes this Fraunhofer diffraction; it's also referred to as farfield diffraction. With this assumption we have, approximately,

$$r = r_0 - x sin\theta,$$

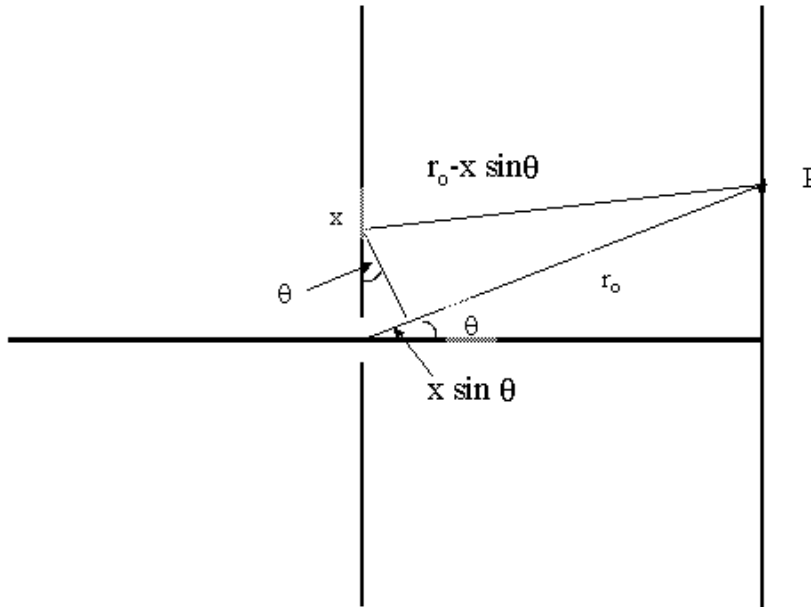where r_0 is the distance between the origin of s to p and θ is the angle between the $z - axis$ and p.

Figure (4.6)

Plug this into the formula for E:

$$E = E_0 e^{2\pi i vt} e^{2\pi i r_0/\lambda} \int_{apertures} e^{-2\pi i x \sin\theta/\lambda} \, dx$$

Drop that constant out front — as you'll see, it won't be important for the rest of our considerations.

We describe the apertures on S by a function $A(x)$, which is zero most of the time (the opaque parts of S) and 1 some of the time (apertures). Thus we can write

$$E \propto \int_{-\infty}^{\infty} A(x) e^{-2\pi i p \sin\theta/\lambda} \, dx$$

It's common to introduce the variable

$$p = \frac{sin\theta}{\lambda}$$

and hence to write

163

$$E \propto \int_{-\infty}^{\infty} A(x) \, e^{-2\pi i p x} \, dx.$$

There you have it. With these approximations (the Fraunhoferapproximations) the electric field (up to a multiplicative constant) is the Fourier transform of the aperture! Note that the variables in the formula are x, a spatial variable, and $p = \sin\theta/\lambda$, in terms of an angle θ. It's the θ that's important, and one always speaks of diffraction "through an angle."

Diffraction by a single slit,Take the case of a single rectangular slit of width$a$, thus described by $A(x) = \Pi_a(x)$. Then the field at $p$is

$$E \, \alpha \, a \, sinc \, ap = a \, sinc \left(\frac{a \, sin\theta}{\lambda} \right).$$

Now, the intensityof the light, which is what we see and what photodetectors register, is proportional to the energy of $E, i.e., to \, |E|^2$. (This is why we dropped the factors $E_0 e^{2\pi i v t} e^{2\pi i r_0/\lambda}$multiplying the integral. They have magnitude 1.) So the diffraction pattern you see from a single slit, those alternating bright and dark bands, is

$$\text{intensity} = a^2 \sin c^2 \left(\frac{a \, sin\theta}{\lambda} \right).$$

Pretty good. The $sinc$ function, or at least its square, live and in color. Just as promised.

We've seen a plot of $\sin c^2$before, and you may very well have seen it, without knowing it, as a plot of the intensity from a single slit diffraction experiment. Here's a plot for

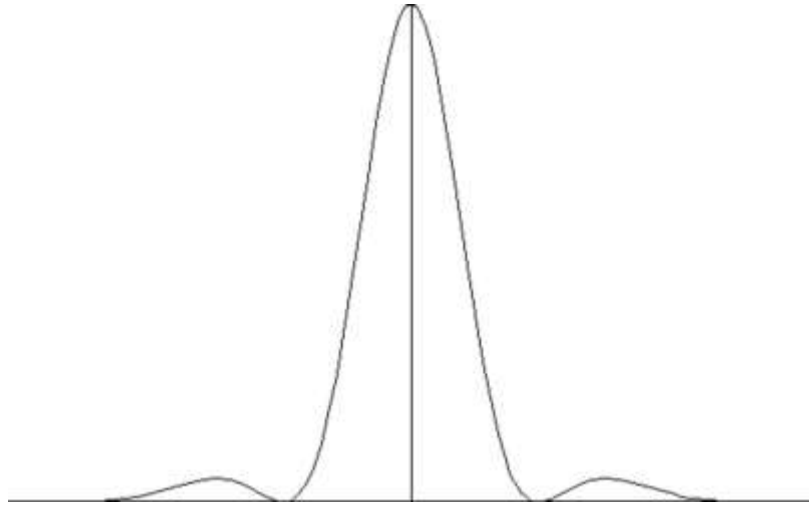$$a = 2, \lambda = 1 \, and \, -\pi/2 \le \theta \le \pi/2:$$

Figure (4.7)

Young's experiment, as mentioned earlier, Thomas Young observed diffraction caused by light passing through two slits. To analyze his experiment using what we've derived we need an expression for the apertures that's convenient for taking the Fourier transform.

Suppose we have two slits, each of width a, centers separated by a distance b. We can model the aperture function by the sum of two shifted rect functions,

$$A(x) = \Pi_a(x - b/2) + \Pi_a(x + b/2).$$

(Like the transfer function of a bandpass filter.) That's fine, but we can also shift the Π_a's by convolving with shifted δ's, as in

$$A(x) = \delta(x - b/2) * \Pi_a(x) + \delta(x + b/2) * \Pi_a x)$$
$$= (\delta(x - b/2) + \delta(x + b/2))$$
$$* \Pi_a(x),$$

and the advantage of writing $A(x)$ in this way is that the convolution theorem applies to help in computing the Fourier transform. Namely,

$$E(p)\alpha(2 \cos \pi bp)(a \ sinc \ ap) = 2a \cos\left(\frac{\pi b sin\theta}{\lambda}\right) sinc\left(\frac{a \ sin\theta}{\lambda}\right)$$

Young saw the intensity, and so would we, which is then

$$intensity = 4a^2 cos^2\left(\frac{\pi b sin\theta}{\lambda}\right) sinc^2\left(\frac{a \ sin\theta}{\lambda}\right)$$

Here's a plot for $a = 2, b = 6, \lambda = 1 \ for \ -\pi/2 \leq \theta \leq \pi/2$:
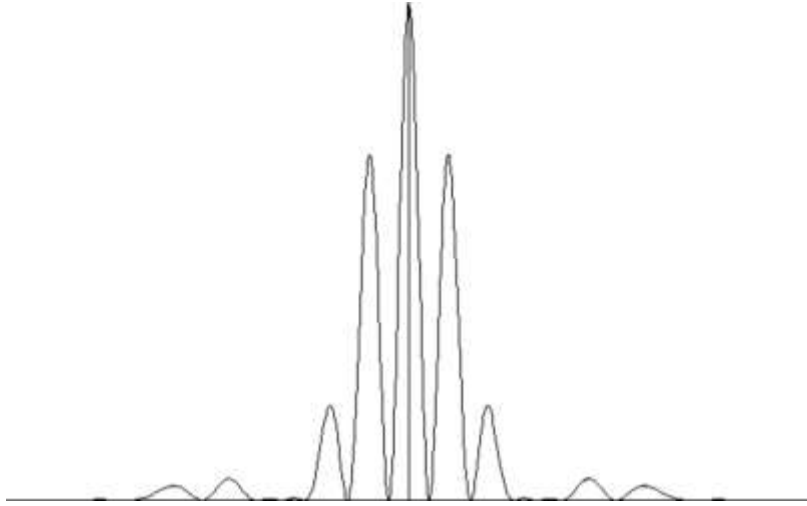
Figure (4.8)

This is quite different from the diffraction pattern for one slit.Diffraction by two point-sources, Say we have two point-sources — the apertures — and that they are at a distance b apart. In this case we can model the apertures by a pair of δ-functions:

$$A(x) \; = \; \delta(x \, - \, b/2) + \; \delta(x + \, b/2).$$

Taking the Fourier transform then gives

$$E(p)\alpha 2 \cos \pi b p = 2 \cos\left(\frac{\pi b \sin\theta}{\lambda}\right).$$

and the intensity as the square magnitude:

$$intensity \; = 4cos^2\left(\frac{\pi b \sin\theta}{\lambda}\right).$$

Here's a plot of this for $b \; = \; 6, \lambda \; = \; 1 \; for \; -\pi/2 \; \leq \; \theta \; \leq \; \pi/2$:
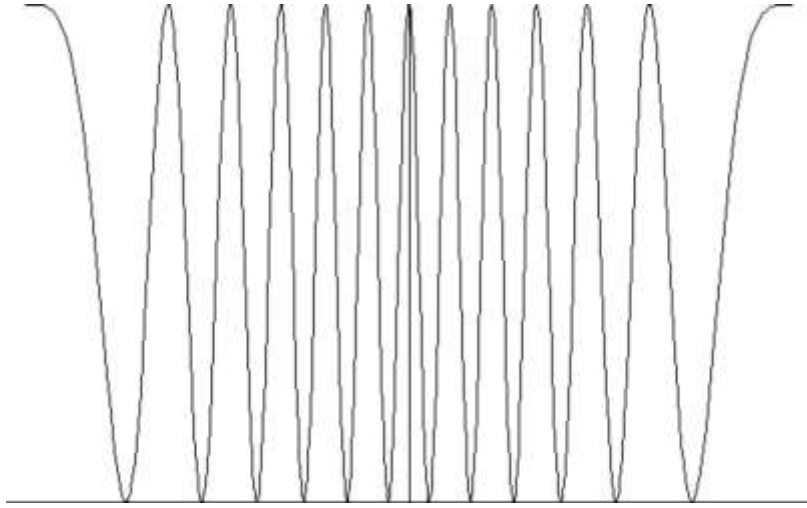
Figure (4.9)

Incidentally, two radiating point sources covers the case of two antennas "transmitting in phase from a single oscillator".

An optical interpretation of $\mathcal{F}\delta = 1$ What if we had light radiating from a single point source? What would the pattern be on the image plane in this circumstance? For a single point source there is no diffraction (a point source, not a circular aperture of some definite radius) and the image plane is illuminated uniformly. Thus the strength of the field is constant on the image plane. On the other hand, if we regard the aperture as δ and plug into the formula we have the Fourier transform of δ,

$$E \, \alpha \int_{-\infty}^{\infty} \delta(x) e^{-2\pi i p x} dx$$

This gives a physical reason why the Fourier transform of δ should be constant (if not 1).

Also note what happens to the intensity as $b \rightarrow 0$ of the diffraction due to two point sources at a distance b.

Physically, we have a single point source (of strength 2) and the formula gives

$$intensity \, = 4cos^2\left(\frac{\pi b sin\theta}{\lambda}\right) \rightarrow 4.$$

References

(1) Bracewell, The Fourier Transform and its Applications, McGraw Hill, 1986.

(2) Brigham, The Fast Fourier Transform, Prentice Hall, 1974.

(3) Dym and H. P. McKean, Fourier Series and Integrals, Academic Press, 1972.

(4) Gray and J. W. Goodman Fourier Transforms, Kluwer, 1995.

(5) Jack D. Gaskill, Linear Systems, Fourier Transforms, and Optics, Wiley, 1978.

(6) James, A Student's Guide to Fourier Transforms, Cambridge, 1995.

(7) K̈orner, Fourier Analysis, Cambridge, 1988.

(8) Nahim, The Science of Radio, 2nd edition, Springer, 2001.

(9) Papoulis, The Fourier Transform and its Applications, McGraw Hill, 1960.

(10) Papoulis, Systems and Transforms With Applications in Optics, Krieger Publishing Company, 1981.

(11) Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw Hill, 1991.

(12) Strichartz, A Guide to Distribution Theory and Fourier Transforms, CRC Press, 1994.

(13) Terras, Harmonic Analysis on Symmetric Spaces and Applications, I, II, Springer Verlag, 1988.

(14) van Loam, Computational Frameworks for the Fast Fourier Transform, SIAM 1992.

(15) W. Briggs and V. Henson, The DFT: An Owner's Manual for the Discrete Fourier Transform, SIAM,1995.