Contents

1 Hilbert space and orthonormal bases 2
  1.1 Norms, inner products and Schwarz inequality . . . . . . . . . . . 2
  1.2 Bessel’s inequality and orthonormal bases . . . . . . . . . . . . . 5
  1.3 Problems on Hilbert space . . . . . . . . . . . . . . . . . . . . . . . 12

2 Generalized functions. (δ functions and all that.) 14
  2.1 Dual spaces . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
  2.2 Problems on Linear Functionals . . . . . . . . . . . . . . . . . . . . 18
  2.3 Generalized functions . . . . . . . . . . . . . . . . . . . . . . . . . 20
  2.4 Derivatives of generalized functions . . . . . . . . . . . . . . . . . 24
  2.5 Problems on derivatives and convergence of distributions. . . . . . . 26
  2.6 Distributions over \( \mathbb{R}^n \) . . . . . . . . . . . . . . . . . 28
  2.7 Poisson’s Equation . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

3 The Fourier Transform 32
  3.1 The Fourier Inversion formula on \( S(\mathbb{R}) \) . . . . . . . . . . . 36
  3.2 Extension of the Fourier transform to \( L^2 \) and other function
      spaces. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
  3.3 The Fourier transform over \( \mathbb{R}^n \) . . . . . . . . . . . . . . . . 43
  3.4 Tempered Distributions . . . . . . . . . . . . . . . . . . . . . . . . 46
  3.5 Problems on the Fourier Transform . . . . . . . . . . . . . . . . . 49
1 Hilbert space and orthonormal bases

1.1 Norms, inner products and Schwarz inequality

Definition. A norm on a real or complex vector space, $V$ is a real valued function, $\| \cdot \|$ on $V$ such that

1. $\|x\| \geq 0$ for all $x$ in $V$.
2. $\|\alpha x\| = |\alpha|\|x\|$ for all scalars $\alpha$ and all $x$ in $V$.
3. $\|x + y\| \leq \|x\| + \|y\|$.

Note that the converse of 1.b follows from 2. because $\|O\| = \|0O\| = 0$.

Examples

1. $V = \mathbb{R}$, $\|x\| = |x|$.
2. $V = \mathbb{C}$, $\|x\| = |x|$.
3. $V = \mathbb{R}^n$ with $\|x\| = \sqrt{\sum_{j=1}^{n} x_j^2}$ when $x = (x_1, \ldots, x_n)$.
4. $V = \mathbb{C}^n$ with $\|x\| = \sqrt{\sum_{j=1}^{n} |x_j|^2}$ when $x = (x_1, \ldots, x_n)$.
5. $V = C([0, 1])$. These are the continuous, complex valued functions on $[0, 1]$.

Here are two norms on $V$.

\[
\|f\|_1 = \int_{0}^{1} |f(t)|dt \tag{1.1} \]

\[
\|f\|_\infty = \sup\{|f(t)| : t \in [0, 1]\} \tag{1.2} \]

Exercise 1.1 Prove that the expression in (1.2) is a norm.

Here is an infinite family of other norms on $\mathbb{R}^n$. Let $1 \leq p < \infty$. Define

\[
\|x\|_p = \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p} \tag{1.3} \]

FACT: These are all norms. And here is yet one more. Define

\[
\|x\|_\infty = \max_{j=1, \ldots, n} |x_j| \tag{1.4} \]
Exercise 1.2  Prove that $\|x\|_1$ and $\|x\|_\infty$ are norms. (Sadly, the proof that $\|x\|_p$ is a norm is not that easy. So don’t even try. We’ll prove the $p = 2$ case later.)

The **unit ball** of a norm $\| \cdot \|$ on a vector space $V$ is

$$B = \{ x \in V : \|x\| \leq 1 \} \quad (1.5)$$

A picture of the unit ball of a norm gives some geometric insight into the nature of the norm. Its especially illuminating to compare norms by comparing their unit balls. Lets consider the family of norms defined in (1.3) and (1.4) on $\mathbb{R}^2$. The boundary of each unit ball is the curve $\|x\| = 1$ and clearly goes through the points $(1,0)$ and $(0,1)$. Its not hard to compute that the unit ball of $\| \cdot \|_\infty$ is a square with sides parallel to the coordinate axes. The unit ball of $\| \cdot \|_2$ is a disk contained in this square, and unit ball of $\| \cdot \|_1$ is a diamond shape thing contained in this disk. Less obvious, but reasonable, is the fact that the other unit balls lie inside the square and contain the diamond. Draw pictures.

**Definition** An **inner product** on a real or complex vector space $V$ is a function on $V \times V$ to the scalars such that

1. $(x,x) \geq 0$ for all $x \in V$ and $(x,x) = 0$ only if $x = 0$.
2. $(ax + by, z) = a(x,z) + b(y,z)$ for all scalars $a, b$ and all vectors $x, y$.
3. $(x,y) = (y,x)$.

**Examples**

1. $V = \mathbb{R}^n$ with $(x,y) = \sum_{j=1}^n x_j y_j$.
2. $V = \mathbb{C}^n$ with $(x,y) = \sum_{j=1}^n x_j \overline{y}_j$.
3. $V = C([0,1])$ with $(f,g) = \int_0^1 f(t) \overline{g(t)} dt$. [It would be good for you to verify yourself that this really is an inner product.]
4. $V = l^2$, which is the standard notation for the set of sequences $x = (a_1, a_2, \ldots)$ such that $\sum_{k=1}^\infty |a_k|^2 < \infty$. Define

$$(x,y) = \sum_{k=1}^\infty a_k \overline{b}_k$$

when $y = (b_1, b_2, \ldots)$. The series entering into this definition converges absolutely because of the identity $|ab| \leq (|a|^2 + |b|^2)/2$. [It would be very good for you to verify that this really defines an inner product on $l^2$.]


Theorem 1.1 (The Schwarz inequality.) In any inner product space

\[ |(x,y)| \leq (x,x)^{1/2} (y,y)^{1/2} \tag{1.6} \]  

Proof: If \( x \) or \( y \) is zero then both sides are 0. So we can assume \( x \neq 0 \) and \( y \neq 0 \). In case the inner product space is complex choose \( \alpha \in \mathbb{C} \) such that \(|\alpha| = 1 \) and \( \alpha (x,y) \) is real. If the inner product space is real just take \( \alpha = 1 \). In either case let \( p(t) = \|\alpha x + ty\|^2 \) for all \( t \in \mathbb{R} \). Then

\[ 0 \leq p(t) = (\alpha x + ty, \alpha x + ty) = \|x\|^2 + t^2 \|y\|^2 + 2t (\alpha x, y) \]

because \((\alpha x, y) + (y, \alpha x) = 2 \text{Re}(\alpha x, y) = 2(\alpha x, y)\). So, by the quadratic formula, the discriminant, \( b^2 - 4ac \leq 0 \). That is, \( 4(\alpha x, y)^2 - 4\|x\|^2 \|y\|^2 \leq 0 \). Thus \( |\alpha (x,y)|^2 \leq \|x\| \|y\| \). Now use \( |\alpha| = 1 \). QED.

We are going to show next that in an inner product space one can always produce a norm from the inner product by means of the definition

\[ \|x\| = \sqrt{(x,x)}. \tag{1.7} \]

Corollary 1.2 Define \( \| \cdot \| \) by (1.7). Then

\[ \|x + y\| \leq \|x\| + \|y\|. \]

Proof: Using (1.6) we find

\[ \|x + y\|^2 = (x + y, x + y) = \|x\|^2 + \|y\|^2 + 2 \text{Re}(x, y) \leq \|x\|^2 + 2 \|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \]

QED.

Corollary 1.3 \( \| \cdot \| \) is a norm.

Proof: Using the previous corollary its easy to verify the three properties in the definition of a norm. QED
1.2 Bessel’s inequality and orthonormal bases

**Definition** An orthonormal sequence in an inner product space is a set \(\{e_1, e_2, \ldots\}\) (which we allow to be finite or infinite) such that

\[
(e_j, e_k) = 0 \text{ if } j \neq k \quad \text{and} \quad 1 \text{ if } j = k
\]

**Lemma 1.4** (Bessel’s inequality) Let \(e_1, e_2, \ldots\) be an orthonormal set in a (real or complex) inner product space \(V\). Then for any \(x \in V\)

\[
\|x\|^2 \geq \sum_{j=1}^{\infty} |(x, e_j)|^2
\]

**Proof:** Let \(a_k = (x, e_k)\). Then, for any integer \(n\),

\[
0 \leq (x - \sum_{k=1}^{n} a_k e_k, x - \sum_{k=1}^{n} a_k e_k) = \|x\|^2 - \sum_{k=1}^{n} |a_k|^2 + \sum_{j,k} \overline{a_j} a_k (e_j, e_k)
\]

\[
= \|x\|^2 - \sum_{k=1}^{n} |a_k|^2 - \sum_{k=1}^{n} |a_k|^2 + \sum_{k=1}^{n} |a_k|^2
\]

\[
= \|x\|^2 - \sum_{k=1}^{n} |a_k|^2
\]

So \(\sum_{k=1}^{n} |a_k|^2 \leq \|x\|^2\) for all \(n\). Now take the limit as \(n \to \infty\). QED

**Definition** In a vector space with a given norm we define

\[
\lim_{n \to \infty} x_n = x
\]

to mean

\[
\lim_{n \to \infty} \|x_n - x\| = 0.
\]

We’re all familiar with the concept of an orthonormal basis in finite dimensions: \(e_1, \ldots, e_n\) is an orthonormal basis of a finite dimensional inner product space \(V\) if

(a) the set \(\{e_1, \ldots, e_n\}\) is orthonormal and
(b) every vector \( x \in V \) is a sum: \( x = \sum_{j=1}^{\infty} a_j e_j \).

If \( V \) is infinite dimensional then we should expect that the concept of orthonormal basis should, similarly, be given by the requirement (a) (as before) and (b) every vector \( x \in V \) is a sum

\[
x = \sum_{j=1}^{\infty} a_j e_j. \tag{1.8} \]

And this is right. Of course writing an infinite sum means, as usual, a limit of finite sums: \( \lim_{n \to \infty} \|x - \sum_{j=1}^{n} a_j e_j\| = 0 \). There is an unfortunate and downright annoying aspect to the equation (1.8), however. We see from Bessel’s inequality that (1.8) implies that \( \sum_{k=1}^{\infty} |a_k|^2 < \infty \).

Suppose that we are given the orthonormal sequence \( \{e_1, e_2, \ldots\} \) and a sequence of (real or complex) numbers \( a_j \) such that \( \sum_{k=1}^{\infty} |a_k|^2 < \infty \). Does the series \( \sum_{k=1}^{\infty} a_k e_k \) converge to some vector in \( V \)? If not can we really say then that we have “coordinatized” \( V \) if we don’t even know which coordinate sequences \( \{a_1, a_2, \ldots\} \) actually correspond to vectors in \( V \) (by the formula (1.8))? Here is an example of how easily convergence of the series \( \sum_{k=1}^{\infty} a_k e_k \) can fail, even when \( \sum_{k=1}^{\infty} |a_k|^2 < \infty \).

Let \( F \) be the subspace of \( l^2 \) consisting of finitely nonzero sequences. Thus \( x \in F \) if \( x = (a_1, \ldots, a_n, 0, 0, 0, \ldots) \). \( F \) is clearly a vector space and the inner product on \( l^2 \) restricts to an inner product on \( F \). Let \( e_1 = (1, 0, 0, 0, \ldots) \), \( e_2 = (0, 1, 0, 0, \ldots) \), etc. The sequence \( \{e_1, e_2, \ldots\} \) is orthonormal. Let \( a_j = 2^{-j} \). Then \( \sum_{j=1}^{\infty} |a_j|^2 < \infty \). Now the sequence of partial sums, \( x_n = \sum_{j=1}^{n} a_j e_j \) converges to the vector \( x = \sum_{j=1}^{\infty} a_j e_j \) in \( l^2 \) (you verify this). But \( x \) is not in \( F \). So there is no vector in \( F \) whose coordinates are the nice sequence \( \{2^{-j}\} \). Of course we caused this trouble by making “holes” in \( l^2 \). These circumstances are analogous to the “holes” in the set \( Q \) of rational numbers. For example the sequence \( s_n = \sum_{k=1}^{n} 1/k! \) is a sequence of rational numbers whose limit is \( e \). But \( e \) is not rational. So there is a hole in \( Q \) at \( e \). Question: Can you immagine how intolerably complicated calculus would be if we had to worry about these holes in \( Q \)? (E.g. \( f'(x) = 0 \) gives the maximum of \( F \) on \( Q \) provided \( x \) is rational!) The same nuisance would arise if we allowed holes when dealing with ON bases. We are going to eliminate holes!

**Definition.** A sequence \( x_1, x_2, \ldots \) in a normed vector space is a *Cauchy sequence* if

\[
\lim_{n,m \to \infty} \|x_n - x_m\| = 0.
\]
That is, for any \( \epsilon > 0 \) there is an integer \( N \) such that \( \| x_n - x_m \| < \epsilon \) whenever \( n \) and \( m \geq N \).

**Remark.** If \( x_n \) converges to a vector \( x \) in any normed space then the sequence is a Cauchy sequence. Proof: Same as proof of Proposition \( \text{PropR3} \) in the Appendix. Just replace \(| \cdot |\) by \( \| \cdot \| \).

**Definition.** A normed vector space is **complete** if every Cauchy sequence in \( V \) has a limit in \( V \). A **Banach** space is a normed vector space which is complete.

**Examples.** The spaces \( V \) in Examples 1.1 to 1.4 are complete. [You’re supposed to know this from first year calculus.]

In Example 1.5 the space \( C([0, 1]) \) is complete in the norm \( \| f \|_\infty \) but not in the norm \( \| f \|_1 \). [See if you can prove both of these statements.]

**Definition.** A **Hilbert space** is an inner product space which is complete in the associated norm (H3).

**Examples.** The Examples 1.6, 1.7, 1.9 are Hilbert spaces. But Example (1.8) is not complete. The proof that Example 1.9 is complete will be given only on popular demand. It is extremely unfortunate that the Example 1.8 is not complete. In order to get a complete space one must throw in with the continuous functions all the functions whose square is integrable. This is such an important example that it gets its own notation.

**Notation.**

\( L^2(0, 1) \) is the set of functions \( f : (0, 1) \to \mathbb{C} \) such that \( \int_0^1 |f(t)|^2 dt < \infty \)

For these functions we define

\[
(f, g) = \int_0^1 f(t) \overline{g(t)} dt
\]

This is an inner product on \( L^2(0, 1) \) (easy to verify) and the associated norm is

\[
\| f \| = \sqrt{\int_0^1 |f(t)|^2 dt}
\]

Of course if we wish to consider square integrable functions on some other set, such as \( R \) we would denote it by \( L^2(R) \).

Just as the real numbers fill in the “holes” in the rational numbers so also one may view \( L^2(0, 1) \) as filling in the “holes” in \( C([0, 1]) \). Here is the definition that makes this notion precise.
Definition. A subset $A$ of a Hilbert space $H$ is called dense if for any $x \in H$ and any $\epsilon > 0$ there is a vector $y$ in $A$ such that $\|x - y\| < \epsilon$. In words: you can get arbitrarily close to any vector in $H$ with vectors in $A$. As we know, the rational numbers are dense in the real numbers. For example the rational number $3.141592650000000$ is pretty close to $\pi$. Similarly, if you cut off the decimal expansion of a real number at the twentieth digit after the decimal point then you have a rational number which is very close to the given real number.

GOOD NEWS: $C([0, 1])$ is dense in $L^2(0, 1)$. You may use this fact whenever you find it convenient. Its often best to prove some formula for an easy to handle dense set first, and then show that it automatically extends to the whole Hilbert space. We’ll see this later in the context of Fourier transforms.

Here is the first important consequence of completeness.

Lemma 1.5 Suppose that $H$ is a Hilbert space and that $\{e_1, e_2, \ldots\}$ is any ON sequence. Let $c_k$ be any sequence of scalars such that $\sum_{k=1}^{\infty} |c_k|^2 < \infty$. Then the series

$$\sum_{k=1}^{\infty} c_k e_k$$

converges to some unique vector $x$ in $H$.

Proof: Let $s_n = \sum_{k=1}^{n} c_k e_k$. We must show that the sequence converges to a vector in $H$. Since we don’t know in advance that there is a limit, $x$, to which the sequence converges we will show instead that the sequence is a Cauchy sequence. Suppose that $n > m$. Then $\|s_n - s_m\|^2 = \|\sum_{k=m+1}^{n} c_k e_k\|^2 = \sum_{k=m+1}^{n} |c_k|^2 \rightarrow 0$ as $m, n \rightarrow \infty$ because the series $\sum_{k=1}^{\infty} |c_k|^2$ converges. Now, because $H$ is assumed to be complete, we know that there exists a vector $x$ in $H$ such that $\lim_{n \rightarrow \infty} s_n = x$. Since limits are unique, $x$ is unique. (You prove this on the way to your next class.) Of course we will write

$$x = \sum_{k=1}^{\infty} c_k e_k,$$

keeping in mind that this means that $x$ is the limit of the finite sums. QED.

Lemma 1.6 In any inner product space the function $x \rightarrow (x, y)$ is continuous for each fixed element $y$. 

8
Proof. If $x_n \to x$ then $|(x_n, y) - (x, y)| = |(x_n - x, y)| \leq \|x_n - x\||y| \to 0$.

QED

Of course $(x, y)$ is also a continuous function of $y$ for each fixed $x$. One can either repeat the preceding proof or just use $(y, x) = (x, y)$. 
Theorem 1.7 Let \( e_1, e_2, \ldots \) be an orthonormal sequence in a (real or complex) Hilbert space \( H \). Then the following are equivalent.

- a. \( e_1, e_2, \ldots \) is a maximal ON set. That is, it is not properly contained in any other ON set.

- b. For every vector \( x \in H \) we have

\[
x = \sum_{k=1}^{\infty} a_k e_k \quad \text{where} \quad a_k = (x, e_k)
\]

- c. For every pair of vectors \( x \) and \( y \) in \( H \) we have

\[
(x, y) = \sum_{k=1}^{\infty} a_k b_k \quad \text{where} \quad a_k = (x, e_k) \quad \text{and} \quad b_k = (y, e_k)
\]

- d. For every vector \( x \) in \( H \) we have

\[
\|x\|^2 = \sum_{k=1}^{\infty} |a_k|^2
\]

Proof: We will show that a. \( \implies \) b. \( \implies \) c. \( \implies \) d. \( \implies \) a.

Assume that a. holds. Let \( x \in H \). By Bessel we have \( \sum_{k=1}^{\infty} |a_k|^2 < \infty \). By Lemma 1.5, \( y = \sum_{k=1}^{\infty} a_k e_k \) exists. But \( (x-y, e_j) = a_j - \lim_{n \to \infty} (\sum_{k=1}^{n} a_k e_k, e_j) = a_j - a_j = 0 \) for all \( j \). If \( x \neq y \) then let \( h = (x-y)/\|x-y\| \). One can now adjoin \( h \) to the original set and obtain a larger ON set. So we must have \( x - y = 0 \). This proves that b. holds.

Assume now that b. holds. Then

\[
(x, y) = \lim_{n \to \infty} (\sum_{j=1}^{n} a_j e_j, y) = \lim_{n \to \infty} \lim_{m \to \infty} (\sum_{j=1}^{n} a_j e_j, \sum_{k=1}^{m} b_k e_k)
\]

\[
= \lim_{n \to \infty} \lim_{m \to \infty} \sum_{j=1}^{\min(n,m)} a_j b_j = \sum_{j=1}^{\infty} a_j b_j.
\]

So c. holds.

Next, assume that c. holds. Put \( y = x \) to derive that d. holds.

Finally, assume that d. holds. If \( e_1, e_2, \ldots \) is not a maximal ON set then there exists a vector \( x \neq 0 \) such that \( (x, e_k) = 0 \) for all \( k \). So the “coordinates”, \( a_k = (x, e_k) \) are all zero. But from d. we see that \( \|x\|^2 = \sum_{k=1}^{\infty} |a_k|^2 = 0 \). So \( x = 0 \). Contradiction. QED.
Now we’re ready for the definition of ON basis.

**Definition.** An ON sequence \( \{e_1, e_2, \ldots \} \) in a Hilbert space \( H \) is an ON basis of \( H \) if condition b. in Theorem \text{thmH.5} holds.

Of course we could have used any of the other conditions in Theorem \text{thmH.5} for the definition of ON basis because they’re equivalent. So why did I use condition b. for the definition? Because surveys of your predecessors show that it’s the most popular.
1.3 Problems on Hilbert space

1. Let \( f_1, f_2, \ldots, f_9 \) be an orthonormal set in \( L^2(0, 1) \). Assume that
   \[ \text{(A)} \quad \int_0^1 6xf_1(x)dx = 2 \quad \text{and} \quad \int_0^1 6xf_2(x)dx = 2\sqrt{2} \]
   What can you say about the value of \( \int_0^1 6xf_5(x)dx \)?
   Give reasons.

2. Let \( u_1(x) = 1/\sqrt{2}, -1 \leq x \leq 1 \) and \( u_2(x) = \sqrt{3}x, -1 \leq x \leq 1. \)
   Suppose that \( f \) and \( g \) are in \( L^2([-1, 1]) \) and \( \|f - g\| \leq 5 \). Let \( a_j = \int_{-1}^1 u_j(x)f(x)dx \) and \( b_j = \int_{-1}^1 u_j(x)g(x)dx \).
   Show that \( \sum_{j=1}^2 |a_j - b_j|^2 \leq 25 \). Cite any theorem you use.

3. Suppose that \( f : [-1, 1] \to \mathbb{R} \) satisfies
   \[ \int_{-1}^1 |f(x)|^2dx = 21 \]
   and
   \[ \int_{-1}^1 f(x)dx = 6. \]
   What can you say about the size of \( \int_{-1}^1 xf(x)dx \)?

4. Suppose that \( u_1, u_2 \) are O.N. vectors in an inner product space \( H \).
   Let \( f \in H \) and assume that
   \[ \|f\|^2 = |a_1|^2 + |a_2|^2 \]
   where \( a_j = (f, u_j) \) for \( j = 1, 2 \).
   Show that \( f = a_1u_1 + a_2u_2. \)
5. The Hermite polynomials are the sequence of polynomials $H_n(x)$ uniquely determined by the properties:

a) $H_n(x)$ is a real polynomial of degree $n$, $n = 0, 1, 2, \ldots$ with positive leading coefficient.

b) The functions $u_n(x) = H_n(x)e^{-x^2/4}$ form an O.N. sequence in $L^2(R)$.

Fact that you may use: If $f$ is in $L^2(R)$ and $\int_{-\infty}^{\infty} f(x)u_n(x)dx = 0$ for $n = 0, 1, 2, \ldots$ then $f = 0$.

Let $c_n = \int_{-\infty}^{\infty} e^{-|x|}u_n(x)dx$.

Evaluate $\sum_{n=0}^{\infty} c_n^2$.

6. Let $\{f_1, f_2, \ldots\}$ be an O.N. set in a Hilbert space $H$. Prove that it is an O.N. basis if and only if the finite linear combinations

$$\sum_{j=1}^{n} a_j f_j \ (n \ finite \ but \ arbitrary)$$

are dense in $H$.

7. Let $\{f_1, f_2, \ldots\}$ be an O.N. set in a Hilbert space $H$. Prove that it is an O.N. basis if and only if Parseval’s equality

$$\|g\|^2 = \sum_{j=1}^{\infty} |(f_j, g)|^2$$

holds for a dense set of $g$ in $H$.

8. Suppose that $\{f_1, f_2, \ldots\}$ is an O.N. basis of a Hilbert space $H$ and that $\{g_1, g_2, \ldots\}$ is an O.N. sequence in $H$. Suppose further that

$$\sum_{j=1}^{\infty} \|g_j - f_j\|^2 < 1.$$

Prove that $\{g_1, g_2, \ldots\}$ is also an O.N. basis.

Hint: Use Theorem 1.1 by supposing that there exists $h \neq 0$ such that $(h, g_j) = 0$ for all $j$. 

13
2 Generalized functions. (δ functions and all that.)

2.1 Dual spaces

The concept of a dual space arises naturally in differential geometry, mechanics and general relativity. And we will need it later to understand generalized functions.

Definition 5.1 Let \( V \) be a real or complex vector space. A function \( L : V \to \text{scalars} \) is called a linear functional if

\[
L(\alpha x + \beta y) = \alpha L(x) + \beta L(y)
\]

for all \( x \) and \( y \) in \( V \) and for all scalars \( \alpha \) and \( \beta \). In other words a linear functional is a linear transformation from \( V \) into \( \mathbb{R} \) (or \( \mathbb{C} \) if \( V \) is a complex vector space.) The dual space to \( V \) is the set, denoted \( V^* \), of all linear functionals on \( V \).

For example the function which is identically 0 is a linear functional. (Check this against the definition.) Moreover if \( L_1 \) and \( L_2 \) are linear functionals then so is the function \( aL_1 + bL_2 \) for any scalars \( a \) and \( b \). (Check this against the definition now!) Therefore \( V^* \) is itself a vector space. Its a new vector space constructed from the old one.

Example 5.2. Denote by \( \mathcal{P}_3 \) the vector space consisting of polynomials of degree less or equal to 3. This is a four dimensional vector space because \( \{1, t, t^2, t^3\} \) constitutes a basis. Here are some linear functionals on this space.

1. \( p \mapsto L_1(p) = p(7) \).
2. \( p \mapsto L_2(p) = \int_0^1 p(t)dt \).
3. \( p \mapsto L_3(p) = \int_0^5 p(t) \sin t dt \).

[It would be best if you, personally, verify that each of these functions on \( \mathcal{P}_3 \) are linear functionals.]

The thing to take away from these examples is that there is no resemblance between \( V \) and \( V^* \): you cannot really “identify” any of these three linear functionals on \( \mathcal{P}_3 \) with elements of \( \mathcal{P}_3 \) itself. RIGHT? \( V^* \) is really a different vector space from \( V \) itself. We have constructed a new vector space from the given one. This being the case, you have to regard the following theorem as remarkable.
Theorem 5.3 If $V$ is an $n$-dimensional vector space then so is $V^*$. 

Proof. Let $e_1, \ldots, e_n$ be any basis of $V$. Then any vector $x \in V$ can be uniquely written

$$x = \sum_{j=1}^{n} a_j e_j \quad (2.1)$$

Uniqueness means that each $a_j$ is a function of $x$. Define

$$L_j(x) = a_j, \quad j = 1, \ldots, n.$$ 

Its straightforward to check that each function $L_j$ is a linear functional. We will show that they form a basis of $V^*$. 

1. They are linearly independent. Proof: Suppose that $M := \sum_{j=1}^{n} c_j L_j = 0$. Then $0 = M(e_k) = \sum_{j=1}^{n} c_j L_j(e_k) = \sum_{j=1}^{n} c_j \delta_{jk} = c_k$. So all the coefficients $c_k$ are zero. Hence the functionals $L_j$ are linearly independent.

2. They span $V^*$. Proof: Let $L$ be any linear functional. Define $c_k = L(e_k)$. Claim: Then $L = \sum_{k=1}^{n} c_k L_k$. You can check this yourself by showing that both sides of this equation agree on each $e_j$ and therefore on all linear combinations of the $e_j$. Thus they agree on all of $V$.

So we have now produced a basis of $V^*$ consisting of $n$ elements. Hence $\dim V^* = n$. QED.

Terminology: 5.4 The basis $L_j$ described in the preceding proof is called the dual basis to the basis $e_1, \ldots, e_n$. It has the nice property that

$$L_j(e_k) = \delta_{jk}$$

Philosophic considerations 5.5. Having chosen the basis $e_1, \ldots, e_n$ of $V$ we see that we automatically get a basis $L_1, \ldots, L_n$ of $V^*$. Since any vector $x$ in $V$ can be written uniquely in the form (5.1) we can now define a vector $L_x$ in $V^*$ by the formula

$$L_x = \sum_{j=1}^{n} a_j L_j$$

In this way we get a map $x \mapsto L_x$ from $V$ onto $V^*$. You can check easily that this map is a) linear, b) one-to-one and c) onto $V^*$. This is, as some people would say, an isomorphism from $V$ onto $V^*$. With the help of this map we could, if we wished, identify $V$ and $V^*$ and even go so far as to say that $V$ and $V^*$ are the “same” space. But there is a catch: A choice of basis has been made in constructing this isomorphism. If Jim goes into one room
and chooses a basis $e_1, \ldots, e_n$ to construct this isomorphism and Jane goes into another room and chooses a basis the chances are that they will choose different bases. Then they will arrive at different isomorphisms. So each one will identify $V$ with $V^*$ in different ways. Jim will say that the vector $x \in V$ corresponds to a certain linear functional $L$ and Jane will say, no, it corresponds to a different linear functional, $M$.

When an isomorphism between two vector spaces depends on someone’s choice of a basis we say that the isomorphism is not natural. If you should nevertheless decide to think of these two vector spaces as the “same” (i.e. identify them) then sooner or later you will run into conceptual and even computational trouble.

But there is an important circumstance in which one really can justify identifying $V$ and $V^*$. (Some readers might recognize the next theorem as “raising and lowering” indices.)

**Theorem 5.6.** Suppose that $V$ is a real finite dimensional vector space and $(\cdot, \cdot)$ is a given inner product on $V$. Then for any linear functional $L$ on $V$ there is a unique vector $y$ in $V$ such that

$$L(x) = (x, y) \quad \text{for all } x \in V.$$ 

Denote by $L_y$ the linear functional determined by $y$ in this way. That is,

$$L_y(x) = (x, y) \quad \text{for all } x \in V. \quad (2.2)$$

Then the map

$$y \mapsto L_y$$

is a one-to-one linear map of $V$ onto $V^*$. (I.e. it is an isomorphism.)

**Proof.** The map $y \mapsto L_y$ is clearly linear. (You better check this. It will be good practice in dealing with these structures.) Moreover this map is one-to-one because if $L_y = 0$ then in particular $L_y(y) = 0$. That is, $(y, y) = 0$. So $y = 0$. Therefore the map $y \mapsto L_y$ is one-to-one. Hence, by the rank theorem, the range of this map has the same dimension as the domain. But if $\dim V = n$ then by Theorem 5.3 $\dim V^* = n$ also. Hence the range is all of $V^*$. QED.

**Moral 5.7.** We know that there are many inner products on any finite dimensional vector space. But given a particular inner product on a real finite dimensional vector space $V$, the preceding theorem provides a natural way to identify $V$ with $V^*$ without making any ad hoc choices of basis. Here is a consequence of this identification that we live with every day.
Derivative versus gradient. Suppose that $V$ is a finite dimensional real vector space and $f$ is a real valued function on $V$. For any point $x$ in $V$ and any vector $v \in V$ define

$$\partial_v f(x) = \frac{df(x + tv)}{dt}|_{t=0}$$

This is the derivative of $f$ in the direction $v$. For example if we choose any basis $e_1, ..., e_n$ of $V$ we may write $x = \sum_{j=1}^n x_j e_j$ and then $f$ is just a function of $n$ real variables, $x_1, ..., x_n$. The chain rule then gives

$$\partial_v f(x) = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x)$$

where of course $v = \sum_{j=1}^n v_j e_j$. This sum is clearly linear in $v$. In other words the map $v \rightarrow \partial_v f(x)$ is, for each $x$, a linear functional on $V$. One often writes $f'(x)$ for this linear functional. That is, $f'(x)v = \partial_v f(x)$. So $f'(x)$ is in $V^*$ for each $x \in V$. Therefore the derivative, $f'$, is a function from $V$ into $V^*$. If there is no natural way to identify $V^*$ with $V$ then this map to $V^*$ is the only object around that captures the notion of derivative that you’re familiar with. But if $V$ has a given inner product, $(\cdot, \cdot)$, then we can identify the linear functional $f'(x)$ ( for each $x$) with an element of $V$. This is the gradient of $f$. That is,

$$\nabla f(x) = f'(x)$$

identified to an element of $V$ by Theorem 5.6. Thus

$$f'(x)v = \partial_v f(x) = (\nabla f(x), v).$$
2.2 Problems on Linear Functionals

Definition. A linear functional on a real or complex vector space $V$ is a scalar valued function $f$ on $V$ such that

i) $f(\alpha x) = \alpha f(x)$ for all scalars $\alpha$ and all $x \in V$.
and ii) $f(x + y) = f(x) + f(y)$ for all $x$ and $y$ in $V$.

Which of the following expressions define linear functionals on the given vector space?

1. $V = \mathbb{R}^3$, $x = (x_1, x_2, x_3)$. Explain why not if you think not.
   a) $f(x) = x_1 + 5x_2$
   b) $f(x) = x_1 + 4$
   c) $f(x) = x_1^2 + 5x_3$
   d) $f(x) = 7$
   e) $f(x) = 0$
   f) $f(x) = \sin x_2$
   g) $f(x) = x_1x_2$
   h) $f(x) = x \cdot u$ where $u$ is a fixed vector and $x \cdot u = \sum_{j=1}^{3} x_ju_j$

2. $V = C([0,1])$ (real valued continuous functions on $[0,1]$).
   a) $F(\varphi) = \int_0^1 \varphi(t)dt$ for $\varphi \in C([0,1])$
   b) $F(\varphi) = \varphi(3/5)$
   c) $F(\varphi) = \varphi(0)$
   d) $F(\varphi) = \varphi(0)^2$
   e) $F(\varphi) = \varphi(0)\varphi(1)$
   f) $F(\varphi) = \int_0^1 \varphi(t) \sin tdt$
   g) $F(\varphi) = \int_0^1 \varphi(t)^2dt$
h) \( F(\varphi) = \int_0^1 \varphi(t)^2 t \, dt \)

i) \( F(\varphi) = \int_0^1 (\varphi(t)/\sqrt{t}) \, dt \)

j) \( F(\varphi) = \int_0^1 (\varphi(t)/t) \, dt \)

k) \( F(\varphi) = \int_0^1 (\sin \varphi(t)) \, dt \)

Explain why not if you think not.

3. Let \( V \) be the vector space of all finitely non–zero real sequences. [A sequence \( x = (x_1, x_2, \ldots) \) is called finitely non–zero if \( \exists N \ni x_k = 0 \) for all \( k \geq N \).] If \( a = (a_1, a_2, a_3, \ldots) \) is an arbitrary sequence of real numbers let

\[ f_a(x) = \sum_{j=1}^{\infty} a_j x_j \quad \text{for} \quad x \in V. \quad (2.3) \]

a) Show that the series converges for each \( x \) in \( V \).

b) Show that \( f_a(x) \) is a linear function of \( x \).

c) Show that every linear functional on \( V \) has the form (1). That is, show that if \( f \) is a linear functional on \( V \) then there exists a sequence, \( a \), such that \( f(x) = f_a(x) \forall x \in V \).

d) Show that the sequence \( a \), in part c) is unique.

4. Denote by \( P_2 \) the space of real valued polynomials of degree less or equal to 2. This is a real vector space of dimension three. (Right?) Define an inner product on \( P_2 \) by

\[ (p, q) = \int_{-1}^{1} p(t)q(t) \, dt. \]

You have already admitted that the function on \( P_2 \) defined by \( L(p) = \int_{0}^{2} p(t) \sin t \, dt \) is a linear functional. But we also know that every linear functional is given uniquely by an element of the space \( P_2 \) with the help of the inner product. Thus there is a unique polynomial \( f \) of degree at most 2 such that

\[ L(p) = (p, f) \quad \text{for all} \quad p \in P_2. \]

Find \( f \).
2.3 Generalized functions

If you want to measure the electric field near some point in space you could put a small charged piece of cork there and measure the force exerted by the field on the cork. In this way you have converted the problem of making an electrical measurement to that of making a mechanical measurement. Of course the force on the cork is a (constant multiple of) an average of the forces at each point of the cork. If \( E(x) \) is the field strength at \( x \) and \( \rho \) is the charge distribution on the cork then the net force on the cork is \( \int \mathbb{R}^3 E(x)\rho(x)d^3x \), in appropriate units. In practice (and even in some theories) it is only these averages that you can measure. For example if you really want to measure the field at a point \( x \) by the preceding method you would have to place a point charge at \( x \). But classical theory shows that the total electric energy of the field produced by a point charge is infinite. The notion of a point charge is therefore at best an idealization. Of course if you know in advance that the electric field is continuous then you can get better and better approximations to the value of \( E(x) \) by using a sequence of smaller and smaller corks. In the classical theory of electromagnetic fields the electric field tends to be continuous and therefore it makes sense to talk about its value at a point. The quantum theory of electromagnetic fields, however, recognizes that there are tremendous fluctuations in the field at very small scales and only the averaged field has a meaning.

The need for talking only about averages shows up also in measurement of temperature. A thermometer is clearly measuring the average temperature over the volume of the little bulb at the bottom. If the temperature varies from point to point and is a continuous function of position then you can, in principle, measure the temperature at a point by using a sequence of smaller and smaller bulbs. Recall however that a typical small bulb will contain on the order of \( 10^{22} \) molecules. In accordance with statistical mechanics the temperature at a point of a system has a really questionable meaning because temperature is a measure of average kinetic energy of a large bunch of molecules. So at an atomic size of scale temperature at a point is meaningless.

A pairing such as \( \int \mathbb{R}^3 E(x)\rho(x)d^3x \), between an “extensive” quantity such as charge (”extensive” means that in a larger volume you have more of the stuff, such as charge, mass, etc.) and an “intensive” quantity, such as the electric field (in a larger volume you don’t have more field) occurs often in physics. The integral is linear in \( \rho \) and defines a linear functional on the
vector space of test charges. The integral is also linear in $E$ for fixed $\rho$. We say that the integral is a bilinear pairing between the extensive quantity $\rho$ and the intensive quantity $E$. Such bilinear pairings between two vector spaces is common. Usually the elements in these dual vector spaces have a physical interpretation, extensive for one and intensive for the other. (Julian Schwinger, in his book “Sources and Fields” emphasizes the duality between sources and fields.)

We are going to develop and use this notion of duality between very smooth functions (test functions) and very “rough” functions (e.g. delta functions.) It has proven to be a great simplifying machinery for understanding partial differential equations as well as Fourier transforms. We are going to apply it to both.

The first step is to understand really smooth functions.

**Test functions**

**Lemma 1.** Let

$$f(x) = \begin{cases} 
e^{-1/x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Then $f$ is infinitely differentiable on the entire real line.

**Proof:** First, recall that $e^t$ grows faster than any polynomial as $t \to +\infty$. That is, $\lim_{t \to +\infty} p(t)e^{-t} = 0$ for any polynomial. Second, you can see by induction that for $x > 0$ the $n$th derivative $d^n e^{-1/x} dx^n = p_n(1/x)e^{-1/x}$ for some polynomial $p_n$. [ Convince yourself with the cases $n = 0, 1, 2$.] Third, you can see easily that all of the derivatives of $f$ exist at any point other than $x = 0$, and is zero to the left of 0. The only question then is what's happening at $x = 0$? Sadly, one must go back to the definition of derivative to answer this. But its not so hard. If we know that the first $n$ derivatives exist at $x = 0$ and are zero there then the first and second comments above show that the $n + 1$st right hand derivative is

$$\lim_{h \downarrow 0} \frac{p_n(1/h)e^{-1/h} - 0}{h - 0} = 0$$

Of course the left hand derivative is clearly zero. So the two sided derivative exists and is zero. This is the basis for an induction proof. Carry out the case $n = 0$ yourself. QED.

**Lemma 2** Let

$$\phi(x) = f(x)f(1-x)$$
where $f$ is the function constructed in Lemma 1. Then $\phi$ is an infinitely differentiable function on $\mathbb{R}$ with support contained in the interval $[0, 1]$.

**Proof.** $\phi$ is clearly infinitely differentiable by the repeated application of the product rule for derivatives. To see that $\phi$ is zero outside the interval $[0, 1]$ draw a picture. QED

**Notation.** $C^\infty_c$ is the standard notation for the set of infinitely differentiable functions with support in a finite interval. One says that these functions have *compact support*.

We have now constructed one (not identically zero) function in $C^\infty_c$. From this function its easy to construct lots more. For example the function $\psi(x) = 3\phi(2x + 7) + 5\phi(4x)^3$ is also in $C^\infty_c$. By scaling and translating the argument of $\phi$ and taking powers one clearly gets lots of such functions. In FACT there are so many such functions that they are dense in $L^2(\mathbb{R})$. [Remember the concept of density form the chapter on Hilbert space?] One of the homework problems sketches how to show that any *continuous* function with compact support is a limit of functions in $C^\infty_c$.

**Notation** The space $C^\infty_c$ arises so often that it customarily is given a special notation:

$$\mathcal{D} \equiv C^\infty_c(\mathbb{R}).$$

The dual space of $\mathcal{D}$ is denoted, as usual, $\mathcal{D}^*$.  

**Terminology** An element in $\mathcal{D}^*$ is called a *distribution* or a *generalized function* (according to taste).

**Examples** 1. Suppose that $f$ is a continuous function on $\mathbb{R}$. Define

$$L_f(\phi) = \int_{-\infty}^{\infty} f(x)\phi(x)dx \text{ for } \phi \in \mathcal{D}. \quad (2.4)$$

Then $L_f$ is a linear functional on $\mathcal{D}$. Notice that even if $f$ increases near $\infty$ (e.g. $f(x) = e^{-x^2}$) the integral makes sense because its really an integral over some (and in fact any) interval that supports $\phi$. So $L_f \in \mathcal{D}^*$.

2. Let

$$L_\delta(\phi) = \phi(0).$$

Then $L_\delta$ is also a linear functional on $\mathcal{D}$. (Clear?) So $L_\delta \in \mathcal{D}^*$. But this example is substantially different from the first one. There is no continuous function $f$, or even discontinuous function $f$, such that $L_\delta = \int_{-\infty}^{\infty} f(x)\phi(x)dx$. 

22
3. It's still OK if the function $f$ in Example 1 is not continuous but has only some mild singularities. For example if $f(x) = 1/|x|^{1/2}$ in $(\frac{\bar{B}}{2.4})$ then the integral still exists for any test function $\phi$. All we need of $f$ is that $\int_a^b |f(x)| \, dx < \infty$ for any finite interval $[a, b]$. In particular the function $f(x) = 1/|x|$ won't work in $(\frac{\bar{B}}{2.4})$. The integral doesn't make sense for an arbitrary test function $\phi$. One says that $f$ is locally integrable if $\int_a^b |f(x)| \, dx < \infty$ for any finite interval $[a, b]$.

The lesson to be drawn from these examples is this: any continuous function $f$ on $\mathbb{R}$ produces a "generalized function" $L_f$, i.e. an element of $\mathcal{D}^*$ by means of the formula $(\frac{\bar{B}}{2.4})$. But not every element of $\mathcal{D}^*$ comes from a continuous function in this way (or even from a discontinuous function), as we see in Example 2. That's why we call the elements of $\mathcal{D}^*$ generalized functions. Neat terminology, huh?

4. There is an important instance that violates the wisdom of Example 3. It's based strongly on cancellation of singularity.

Lemma Let $\phi \in \mathcal{D}$. Then

$$P(\frac{1}{x})(\phi) \equiv \lim_{a \downarrow 0} \int_{|x| > a} \frac{\phi(x)}{x} \, dx \quad (2.5)$$

exists and is equal to

$$\int_{|x| > 1} \frac{\phi(x)}{x} \, dx + \int_{-1}^{1} \frac{\phi(x) - \phi(0)}{x} \, dx \quad (2.6)$$

Proof: If $0 < a < b$ then

$$\int_{|x| > a} \frac{\phi(x)}{x} \, dx = \int_{|x| > b} \frac{\phi(x)}{x} \, dx + \int_{a < |x| \leq b} \frac{\phi(x) - \phi(0)}{x} \, dx \quad (2.7)$$

because $\int_{a < |x| \leq b} \frac{1}{x} \, dx = 0$. The integrand in the last term in $(\frac{\bar{B}}{2.7})$ is bounded on $\mathbb{R}$ by the mean value theorem. So we can let $a \downarrow 0$ and get a limit. This also proves the validity of the representation $(\frac{\bar{B}}{2.6})$ of this limit. QED

The generalized function $P(\frac{1}{x})$ is called the Principal Part of $1/x$. It arises in many contexts. We will see it coming up later in the Feynman propagator.
2.4 Derivatives of generalized functions

**Definition** Let \( T \in \mathcal{D}^* \). The derivative of \( T \) is the element \( T' \) of \( \mathcal{D}^* \) given by

\[
T'(\phi) = - T(\phi') \quad \text{for all } \phi \in \mathcal{D}.
\]  

(2.8) \( \text{(D1)} \)

What does this definition have to do with our well known notion of derivative of a function? First lets observe that at least \( T' \) is indeed a well defined linear functional on \( \mathcal{D} \). The reason is that for any \( \phi \in \mathcal{D} \) the function \( \phi' \) is again in \( \mathcal{D} \) so the right side of (2.8) makes sense. The linearity of \( T' \) is clear.

Right? To understand why this definition is justified consider the example \( T = \mathcal{L}_f \) spelled out in (2.4). Suppose that \( f \) is actually differentiable in the classical sense. (i.e. in the sense that you grew up with.) Then

\[
\mathcal{L}_f'(\phi) = \int_{-\infty}^{\infty} f'(x)\phi(x)dx
\]

\[= \int_a^b f'(x)\phi(x)dx\]

where \( a \) and \( b \) are chosen so that \( \phi = 0 \) off \((a,b)\)

\[= f(x)\phi(x)|_a^b - \int_a^b f(x)\phi'(x)dx\]

\[= - \int_a^b f(x)\phi'(x)dx\]

\[= - \int_{-\infty}^{\infty} f(x)\phi'(x)dx\]

\[= -\mathcal{L}_f(\phi')\]

So

\[
\mathcal{L}_f'(\phi) = -\mathcal{L}_f(\phi').
\]  

(2.9) \( \text{(D3)} \)

Stare at (2.8) and (2.9). Do you see now why (2.8) is a justifiable definition of derivative of a generalized function \( T \)? That’s right. It agrees with the classical notion of derivative when \( T = \mathcal{L}_f \) and \( f \) is itself differentiable!!! But (2.8) has a well defined meaning even when its not of the form \( \mathcal{L}_f \) for some differentiable function \( f \). Lets see what the definition (2.8) gives when \( T = \mathcal{L}_H \) and \( H \) is the non-differentiable function given by

\[
H(x) = \begin{cases} 
1, & x \geq 0 \\
0, & x < 0 
\end{cases}
\]

(2.10)
In this case we have, USING (2.8),

\[ T'(\phi) = -T(\phi') \quad (2.11) \]

\[ = -L_H(\phi') \quad (2.12) \]

\[ = -\int_{-\infty}^{0} \phi'(x)dx \quad (2.13) \]

\[ = \phi(0) \quad (2.14) \]

by the fundamental theorem of calculus. So we have

\[ (L_H)' = L_\delta. \quad (2.15) \]

You could say (flippantly) that we have now differentiated a non-differentiable function, \( H \) and found \( H' = \delta \). In truth the perfectly meaningful equation (2.15) is often written as \( H' = \delta \). But what does this equation mean? It means (2.15). For the next two weeks we will regard it as immoral to write the equation \( H' = \delta \).

Example: \( (L_\delta)'(\phi) = -\phi'(0) \)

We now have a notion of derivative of a generalized function. Lets end with one more definition.

**Definition** A sequence \( T_n \) of generalized functions converges to a generalized function \( T \) if the sequence of numbers \( T_n(\phi) \) converges to \( T(\phi) \) for each \( \phi \in D \).
2.5 Problems on derivatives and convergence of distributions.

1. Define 
\[ L(\phi) = \int_{-\infty}^{\infty} |x|\phi(x) \, dx \quad \text{for } \phi \in C^\infty_c(\mathbb{R}). \]

Using the definition 
\[ T'(\phi) = -T(\phi'), \]
compute the first four derivatives of \( L \); that is, compute \( T', ..., T^{(4)} \).

Hint #1. Use the definition of derivative of a distribution.

Hint #2. Use the definition four times to compute the four derivatives.

2. Suppose that \( f : \mathbb{R} \to \mathbb{R} \) is continuous (but not necessarily differentiable.)
Let 
\[ u(x, t) = f(x - ct). \]
Show that \( u \) is a solution to the wave equation 
\[ \frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \]
in the distribution sense (sometimes called the weak sense.)

Nota Bene: Since \( f \) is not necessarily differentiable you cannot use \( f' \) in the classical sense.

3. Does \( \sum_{n=1}^{\infty} \delta(x - n) \) converge in the distribution sense?

4. Let \( p_t(x) = (2\pi t)^{-1/2} e^{-x^2/(2t)} \). Find the following limits if they exist.
   a. the pointwise limit as \( t \downarrow 0 \).
   b. the \( L^2(\mathbb{R}) \) limit.
   c. the limit in \( \mathcal{D}^* \).
You may use the following FACT that will be proved later
\[ \int_{-\infty}^{\infty} p_t(x) \, dx = 1 \quad \forall t > 0. \]

5. Let \( \phi \) be the function constructed in Lemma 2, except multiply it by a positive constant such that
\[ \int_{\mathbb{R}} \phi(x) \, dx = 1 \]
Such a constant can be found because the original $\phi$ is nonnegative and has a strictly positive integral. Then let

$$\phi_n(x) = n\phi(nx)$$

Our goal is to show that $\phi_n$ converges in some sense to a delta function.

a. Show that $\int_{\mathbb{R}} \phi_n(x) dx = 1$ for all positive integers $n$.

b. Suppose that $g$ is a continuous function on $\mathbb{R}$. Show that

$$\lim_{n \to \infty} \int_{\mathbb{R}} \phi_n(x) g(x) dx = g(0).$$

c. Use part b. and the definition of convergence of distributions to show that

$$L_{\phi_n} \text{ converges in the weak sense to } L_\delta.$$  

6. Prove that, for all $\phi \in \mathcal{D}$,

$$\lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty} \frac{\phi(x)}{x + i\epsilon} dx = P\left(\frac{1}{x}\right)(\phi) - i\pi \phi(0) \quad (2.16)$$

This is often written as

$$\lim_{\epsilon \downarrow 0} \frac{1}{x + i\epsilon} = P\left(\frac{1}{x}\right) - i\pi \delta \text{ weak sense} \quad (2.17)$$

Hint: Review the proof of the lemma at the end of Section 3.3.
2.6 Distributions over $\mathbb{R}^n$

The extension of the one dimensional notion of distribution to an $n$ dimensional distribution is straightforward. Denote by $D$ the vector space of infinitely differentiable functions on $\mathbb{R}^n$ which are zero outside of a large cube (that can depend on the function.) This space is sometimes denoted $C^\infty_c(\mathbb{R}^n)$. There are plenty of such functions. For example if $\phi_1, \ldots, \phi_n$ are each in $C^\infty_c(\mathbb{R})$ then the function $\psi(x_1, \ldots, x_n) = \phi(x_1) \cdots \phi(x_n)$ is in $C^\infty_c(\mathbb{R}^n)$. (What cube could you use?) So is any finite linear combination of these products. A distribution over $\mathbb{R}^n$ is defined as a linear functional on $D$. Of course we can make up examples of such $n$-dimensional distributions similar to the ones we already know in one dimension: Let

$$L_f(\phi) = \int_{\mathbb{R}^n} f(x)\phi(x)d^n x \quad (2.18)$$

As long as $|f|$ has a finite integral over every cube this expression makes sense and defines a linear functional on $D$, just as in one dimension. We also have the $n$-dimensional $\delta$ “function” defined by

$$L_\delta(\phi) = \phi(0). \quad (2.19)$$

The difference from one dimension shows up when we consider differentiation. We now have partial derivatives. Here is the definition of partial derivative (as if you couldn’t guess).

$$\frac{\partial T}{\partial x_k}(\phi) = -T(\partial \phi/\partial x_k). \quad (2.20)$$

When $n = 3$ every distribution has a very intuitive interpretation as an arrangement of charges, dipoles, quadrupoles, etc. I’m going to explain this interpretation in class in more detail. It gives physical meaning to every element of $D^*$!!!
2.7 Poisson’s Equation

We will write as usual $r = |x|$ in $\mathbb{R}^3$.

**Theorem 2.1**

$$\Delta \frac{1}{r} = -4\pi \delta.$$ \hspace{1cm} (2.21)

in the distribution sense. That is,

$$\Delta L_{1/r} = -4\pi L_\delta$$

We will break the proof up into several small steps.

**Lemma 2.2** At $r \neq 0$

$$\Delta(1/r) = 0$$

**Proof.** $\partial (1/r)/\partial x = -x/r^3$ and $\partial^2 (1/r)/\partial x^2 = -1/r^3 + 3x^2/r^5$. So

$$\Delta(1/r) = -3/r^3 + 3x^2 + y^2 + z^2/r^5 = -3/r^3 + 3/r^3 = 0.$$ QED.

In view of this lemma you can see that we have only to deal now with the singularity at $r = 0$. Our notion of weak derivative is just right for doing this.

The trick is to avoid the singularity until after one does some clever integration by parts (in the form of the divergence theorem). In case you forgot your vector calculus identities a self contained review is at the end of this section. I want to warn you that this is not the kind of proof that you are likely to invent yourself. But the techniques are so frequently occurring that there is some virtue in following it through at least once in one’s life.

**Lemma 2.3** Let $\phi \in \mathcal{D}$. Then

$$\int_{\mathbb{R}^3} (1/r) \Delta \phi(x) dx = \lim_{\epsilon \to 0} \int_{r \geq \epsilon} (1/r) \Delta \phi dx$$
Proof: The difference between the left and the right sides before taking the
limit is at most (use spherical coordinates in the next step)

\[ \left| \int_{r \leq \epsilon} (1/r) \Delta \phi d^3x \right| \leq \max_{x \in \mathbb{R}^3} |\Delta \phi(x)| \int_{r \leq \epsilon} (1/r) d^3x \]

\[ = \max_{x \in \mathbb{R}^3} |\Delta \phi(x)| 2\pi \epsilon^2 \rightarrow 0 \]

QED.

Before really getting down to business let’s apply the definitions.

\[
\Delta T_{1/r}(\phi) = \sum_{j=1}^{3} (\partial^2 / \partial x_j^2) T_{1/r}(\phi) 
\]

\[= -\sum_{j=1}^{3} (\partial / \partial x_j) T_{1/r}(\partial \phi / \partial x_j) \]

\[= T_{1/r}(\Delta \phi) \]

\[= \int_{\mathbb{R}^3} (1/r) \Delta \phi(x) d^3x \]

\[= \lim_{\epsilon \rightarrow 0} \int_{r \geq \epsilon} (1/r) \Delta \phi(x) d^3x. \]

So what we really need to do is show that this limit is \(-4\pi \phi(0)\). To this end we are going to apply some standard integration by parts identities in the “OK” region \(r \geq \epsilon\).

\[C_\epsilon := \int_{r \geq \epsilon} (1/r) \Delta \phi(x) d^3x \]

\[= \int_{r \geq \epsilon} \nabla \cdot \left( \frac{1}{r} \nabla \phi - \phi \nabla \left( \frac{1}{r} \right) \right) d^3x \quad \text{by identity (P8.39)} \]

\[= \int_{r = \epsilon} \frac{1}{r} \nabla \phi \cdot \mathbf{n} - \phi \left( \frac{1}{r} \right) \cdot \mathbf{n} dA \quad \text{by the divergence theorem} \]

where \(\mathbf{n}\) is the unit normal pointing toward the origin. The other boundary term in this integration by parts identity is zero because we can take it over a sphere so large that \(\phi\) is zero on and outside it.
Now

\[ \int_{r=\epsilon} \frac{1}{r} (\nabla \phi \cdot \mathbf{n}) dA = \frac{1}{\epsilon} \int_{r=\epsilon} (\nabla \phi \cdot \mathbf{n} dA) \]

\[ \leq \frac{1}{\epsilon} (\max |\nabla \phi|) 4\pi \epsilon^2 \]

\[ \rightarrow 0 \]  

as \( \epsilon \downarrow 0 \). This gets rid of one of the terms in \( C_\epsilon \) in the limit. For the other one just note that \( (\nabla \frac{1}{r}) \cdot \mathbf{n} = -\partial(1/r)/\partial r = 1/r^2 \). So

\[ -\int_{r=\epsilon} \phi(\nabla \frac{1}{r}) \cdot \mathbf{n} dA = -\frac{1}{\epsilon^2} \int_{r=\epsilon} \phi(x) dA \]

\[ = -\frac{1}{\epsilon^2} \int_{r=\epsilon} \phi(0) dA - \frac{1}{\epsilon^2} \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \]

\[ = -4\pi \phi(0) - \frac{1}{\epsilon^2} \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \]  

Only one more term to get rid of!

\[ \frac{1}{\epsilon^2} \int_{r=\epsilon} (\phi(x) - \phi(0)) dA \leq \max |\phi(x) - \phi(0)| \cdot 4\pi \rightarrow 0 \]

because \( \phi \) is continuous at \( x = 0 \). This proves \((P1)2.21\).

**Vector calculus identities.**

If \( f \) is a real valued function and \( G \) is a vector field, both defined on some region in \( \mathbb{R}^3 \) then

\[ \nabla \cdot (fG) = (\nabla f) \cdot G + f \nabla \cdot G \]  

**Application #1.** Take \( f = 1/r \) and \( G = \nabla \phi \). Then we get

\[ \nabla \cdot \left( \frac{1}{r} \nabla \phi \right) = \left( \nabla \frac{1}{r} \right) \cdot \nabla \phi + \frac{1}{r} \Delta \phi \]  

wherever \( r \neq 0 \). \((P6)2.37\)

**Application #2.** Take \( f = \phi \) and \( G = \nabla \frac{1}{r} \). Then we get

\[ \nabla \cdot (\phi \nabla \frac{1}{r}) = (\nabla \phi) \cdot (\nabla \frac{1}{r}) + \phi \Delta \frac{1}{r} \]  

wherever \( r \neq 0 \). \((P7)2.38\)

But \( \Delta \frac{1}{r} = 0 \) wherever \( r \neq 0 \). So subtracting \((P7)2.38\) from \((P6)2.37\) we find

\[ \frac{1}{r} \Delta \phi = \nabla \cdot \left( \frac{1}{r} \nabla \phi - \phi \nabla \frac{1}{r} \right) \]  

wherever \( r \neq 0 \). \((P8)2.39\)

This is the identity we need in the proof of \((P1)2.21\).
3 The Fourier Transform

The Fourier transform of a complex valued function on the line is

\[ \hat{f}(\xi) = \int_{-\infty}^{\infty} e^{i\xi x} f(x) dx \]  \hspace{1cm} (3.1) \hspace{1cm} F1

Here \( \xi \) runs over \( \mathbb{R} \). The most useful aspect of this transform of functions is that it interchanges differentiation and multiplication. Thus if you differentiate under the integral sign you get

\[ \frac{d}{d\xi} \hat{f}(\xi) = \int_{-\infty}^{\infty} e^{i\xi x} \{ixf(x)\} dx. \]  \hspace{1cm} (3.2) \hspace{1cm} F2

So

\[ (\text{Fourier transform of}\{ixf(x)\})(\xi) = \frac{d}{d\xi} \hat{f}(\xi). \]  \hspace{1cm} (3.3)

And an integration by parts (never mind the boundary terms) clearly gives

\[ -i\xi \hat{f}(\xi) = \int_{-\infty}^{\infty} e^{i\xi x} f'(x) dx. \]  \hspace{1cm} (3.4) \hspace{1cm} F3

So

\[ \hat{f}'(\xi) = -i\xi \hat{f}(\xi) \]  \hspace{1cm} (3.5)

We will see later that these formulas allow one to solve some partial differential equations. Moreover in quantum mechanics these two formulas amount to the statement that the Fourier transform interchanges \( P \) and \( Q \) (momentum and position operators.)

But the usefulness of these formulas depends crucially on the fact that one can also transform back and recover \( f \) from \( \hat{f} \). To this end there is an inversion formula that does the job. Our goal is to establish the most useful properties of the Fourier transform and in particular to derive the inversion formula and show how to use it to solve PDEs.

To begin with we must understand how to give honest meaning to the formula (3.1). Since the integral is over an infinite interval there is a convergence question right away. Suppose that

\[ \int_{-\infty}^{\infty} |f(x)| dx < \infty. \]  \hspace{1cm} (3.6) \hspace{1cm} F4

32
We will write \( f \in L^1 \) if (3.6) holds. If \( f \in L^1 \) then there is no problem with the existence of the integral in (3.1) because \( \lim_{a \to \infty} \int_a^b e^{i\xi x} f(x) dx \) exists. 

[Proof: \( |(\int_a^b - \int_b^c) e^{i\xi x} f(x) dx| \leq \int_{a \leq |x| \leq b} |f(x)| dx \to 0 \) as \( a \leq b \to \infty \). Use the Cauchy convergence criterion now.]

Of course even if \( f \in L^1 \) it can happen that \( f' \) is not in \( L^1 \) and/or that \( xf(x) \) is not in \( L^1 \). This is a nuisance in dealing with the identities (3.2) and (3.4). We are going to restrict our attention for a while to a class of functions that will make these issues easy to deal with.

**Definition.** A function \( f \) on \( \mathbb{R} \) is said to be **rapidly decreasing** if

\[
|x^n f(x)| \leq M_n, \quad n = 0, 1, 2, \ldots
\]

(3.7)

for some real numbers \( M_n \). In words: \( x^n f(x) \) is bounded on \( \mathbb{R} \) for each \( n \).

**Examples**
1. \( e^{-x^2} \) is rapidly decreasing.
2. \( \frac{1}{x^2+1} \) is not rapidly decreasing because (3.7) only holds for \( n = 0, 1, 2 \) but not for \( n = 3 \) or more.
3. If \( f \) is rapidly decreasing then so is \( x^5 f(x) \) because \( x^n x^5 f(x) = x^{n+5} f(x) \) which is bounded in accordance with (3.7). Just replace \( n \) by \( n + 5 \) in (3.7). Since any finite linear combination of rapidly decreasing functions is also rapidly decreasing we see that \( p(x)f(x) \) is rapidly decreasing for any polynomial \( p \) if \( f \) is rapidly decreasing.
4. So by examples 1. and 3. we see that \( p(x)e^{-x^2} \) is rapidly decreasing for any polynomial \( p \).
5. Any function in \( C_c^\infty(\mathbb{R}) \) is rapidly decreasing.
6. Summary: The space of rapidly decreasing functions is a vector space and is closed under multiplication by any polynomial. And besides, there are lots of these functions.

**Lemma** Any rapidly decreasing function is in \( L^1 \).

**Proof:** Apply (3.7) for \( n = 0 \) and \( n = 2 \) to conclude that

\[
|(1 + x^2) f(x)| \leq M
\]

for some real number \( M \). So

\[
\int_{-\infty}^{\infty} |f(x)| dx \leq M \int_{-\infty}^{\infty} \frac{1}{x^2 + 1} dx < \infty.
\]

QED

Using only rapidly decreasing functions in (3.2) will allow us not to have to worry about whether \( f \) and \( xf(x) \) are both in \( L^1 \). Neat, huh?
But to use (3.4) we still need to worry about whether \( f' \) is in \( L^1 \). So we are going to restrict our attention for a while, even further, to a class of functions that makes both (3.2) and (3.4) easy to deal with.

**Notation:** \( S \) will denote the set of \( C^\infty \) functions on \( \mathbb{R} \) such that \( f \) and each of its derivatives is rapidly decreasing.

**Examples**

7. \( e^{-x^2} \) is infinitely differentiable and each derivative is just a polynomial times \( e^{-x^2} \). We've already seen that these functions are rapidly decreasing. So the function \( e^{-x^2} \) is in \( S \).

8. \( \frac{1}{x^2+1} \) is infinitely differentiable but is not rapidly decreasing. So this function is not in \( S \).

9. Now here is a real nice thing about this space \( S \). If \( f \in S \) then any polynomial, \( p \), times any derivative of \( f \) is again in \( S \). CHECK THIS against the definitions! I know this may seem too good to believe. But we do know that there are lots of functions in \( S \). All of \( C^\infty_c \) is contained in \( S \). And besides there are more, as we saw in Example 7.

**STATUS:** If \( f \in S \) then \( f \) and all of its derivatives are in \( L^1 \). So both formulas (3.2) and (3.4) make sense. Moreover they are both correct because the boundary terms that we ignored in deriving (3.4) are indeed zero, since these functions go to zero so quickly at \( \infty \). (Check this at your leisure!)

In truth, here is the reason that the space \( S \) is so great.

**Invariance Theorem.** If \( f \in S \) then \( \hat{f} \) \( \in S \).

**Proof:** First notice that for any function \( f \) in \( L^1 \) we have the bound

\[
|\hat{f}(\xi)| = |\int_{-\infty}^{\infty} e^{ix\xi} f(x) dx| \leq \int_{-\infty}^{\infty} |f(x)| dx = \|f\|_1.
\]

Since the right side doesn't depend on \( \xi \) \( \hat{f} \) is bounded.

Now suppose that \( f \in S \). Then so is \( f', f'', \) etc. So \( f', f'' \) etc. are all in \( L^1 \). By (3.4) it now follows that \( \xi^n \hat{f}(\xi) \) is bounded for each \( n = 0, 1, 2,... \). So \( \hat{f} \) is rapidly decreasing. But \( d\hat{f}(\xi)/d\xi \) is the Fourier transform of \( ix f(x) \) which we have seen is also in \( S \). So \( df(\xi)/d\xi \) is also rapidly decreasing. And so on. QED.
Here is what the inversion formula will look like

\[ f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iy\xi} \hat{f}(\xi) d\xi. \]  

(3.11)  

Since we already know that \( \hat{f} \) is in \( \mathcal{S} \) we know that the right hand side of (3.11) makes sense. Contrast this with the example that you worked out in the homework: Take \( f(x) = 1 \) if \( |x| \leq 1 \) and \( f = 0 \) otherwise. Then \( f \in L^1 \) so \( \hat{f} \) makes sense. But \( \hat{f} \) itself decreases so slowly at \( \infty \) that its not in \( L^1 \). So (3.11) doesn’t make sense. Aren’t you glad that we are focusing on functions in \( \mathcal{S} \)?

**Notation:**

\[ \tilde{g}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iy\xi} g(\xi) d\xi \]  

(3.12)  

**Inversion Theorem.** For \( f \in \mathcal{S} \) equation (3.11) holds.

In other words : \( \check{\hat{f}} = f \).

We are going to spend the next few pages proving this formula. (Not because I fear that you might not trust me, but because the proof derives some very useful identities along the way.)
3.1 The Fourier Inversion formula on $S(R)$.

To begin, here are some explicit computations of some important Fourier transforms.

**Gaussian identities**

Let

$$p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} \quad (3.13) \quad \text{F10}$$

Then we have the following three identities.

$$\int_{-\infty}^{\infty} p_t(x)dx = 1 \quad (3.14) \quad \text{F11}$$

$$\hat{p}_t(\xi) = e^{-t\xi^2/2} \quad (3.15) \quad \text{F12}$$

$$(e^{-\frac{i}{2}\xi^2})(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} = p_t(x). \quad (3.16) \quad \text{F13}$$

**Proof of (3.14)** [Sneaky use of polar coordinates.]

$$\left\{ \int_{-\infty}^{\infty} p_t(x)dx \right\}^2 = \int_{\mathbb{R}^2} p_t(x)p_t(y)dxdy \quad (3.17)$$

$$= \frac{1}{2\pi t} \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2t}}dxdy \quad (3.18)$$

$$= \frac{1}{2\pi t} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2t}rdrd\theta \quad (3.19)$$

$$= 1 \quad (3.20)$$  

(Use the substitution $s = r^2/(2t)$ in the last step.)

**Proof of (3.15)**

To compute $\hat{p}_t(\xi)$ we need first to multiply $p_t(x)$ by $e^{ix\xi}$ before integration. The exponent is a quadratic function of $x$ for which we can complete the square thus:

$$-\frac{x^2}{2t} + ix\xi = -\frac{1}{2t}(x - it\xi)^2 - t\xi^2/2$$

Hence

$$\hat{p}_t(\xi) = e^{-t\xi^2/2} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{1}{2t}(x-it\xi)^2} dx \quad (3.21) \quad \text{F16}$$
The coefficient of $e^{-t\xi^2/2}$ looks “just like” $\int_{-\infty}^\infty p_t(x)\,dx$ which is one. This would prove (3.15). But not so fast. $x$ has been translated by the imaginary number $it\xi$. You can’t just translate the argument to get rid of it. Here is how to get rid of it. Let $a = -t\xi$. and suppose $a > 0$ just so that I can draw pictures verbally. Take a large positive number $R$ and consider the rectangular contour in the complex plane which starts at $-R$ on the real axis, moves along the real axis to $R$ and then moves vertically up to the point $R + ia$. From there move horizontally, west to $-R + ia$ and then south back to the point $-R$. Since $e^{-x^2/(2t)}$ is an entire function, the integral of this function around this rectangular contour is zero. Otherwise put, $\int_R^{-R}e^{(x+ia)^2/(2t)} = \int_R^{-R}e^{-x^2/(2t)}\,dx$ plus a little contribution from the vertical sides of the rectangle. On the right end the integral is at most $a$ times the maximum value of the integrand on that segment. But $|e^{-(R+iy)^2/(2t)}| = e^{(-R^2+y^2)/(2t)}$, which goes to zero quite quickly for $0 \leq y \leq a$ as $R \to \infty$. So the integrals along the end segments go to zero as $R \to \infty$. So it is indeed true that the coefficient of $e^{-t\xi^2}$ in (3.21) equals $\int_{-\infty}^\infty p_t(x)\,dx$, which is one. QED

Proof of (3.16) $e^{-t\xi^2/2} = \sqrt{\frac{2\pi}{t}}p_{1/t}(\xi)$ by (3.13). Since $p_{1/t}$ is even we have $\hat{e^{-t\xi^2/2}} = (1/2\pi)\int e^{-x^2/(2t)}\,dx$ by (3.15). QED

Definition 3.1 The convolution of two functions $f$ and $g$ on $\mathbb{R}$ is given by

$$ (f \ast g)(x) = \int_R f(x-y)g(y)\,dy $$

(3.22)

Lemma 3.2 Let $g$ be a bounded continuous function on $\mathbb{R}$. Then for each $x$

$$ \lim_{t\to0}(p_t \ast g)(x) = g(x). $$

Proof: Since $\int_{-\infty}^\infty p_t(x)\,dx = 1$ we have

$$ (p_t \ast g)(x) - g(x) = (g \ast p_t)(x) - g(x) $$

(3.23)

$$ = \int g(x-y)p_t(y)\,dy - g(x) $$

(3.24)

$$ = \int_{-\infty}^\infty (g(x-y) - g(x))p_t(y)\,dy. $$

(3.25)
Given $\varepsilon > 0$, $\exists \delta > 0 \ni |g(x - y) - g(x)| < \varepsilon$ if $|y| < \delta$.

$$\therefore |p_t * g(x) - g(x)| \leq \int_{-\delta}^{\delta} |g(x - y) - g(x)| p_t(y) dy$$

$$+ \int_{|y| > \delta} |g(x - y) - g(x)| p_t(y) dy$$

$$\leq \varepsilon \int_{-\delta}^{\delta} p_t(y) dy + 2|g|_{\infty} \int_{|y| > \delta} p_t(y) dy$$

$$\leq \varepsilon + 2|g|_{\infty} \int_{|y| \geq \delta} p_t(y) dy. \quad (3.28)$$

$$\therefore \lim_{t \downarrow 0} |(p_t * g)(x) - g(x)| \leq \varepsilon + 2|g|_{\infty} \lim_{t \downarrow 0} \int_{|y| \geq \delta} p_t(y) dy.$$

But

$$\int_{y \geq \delta} p_t(y) dy = \frac{2}{\sqrt{2\pi t}} \int_{\delta}^{\infty} e^{-y^2/2t} dy \leq \frac{2}{\sqrt{2\pi t}} \int_{\delta}^{\infty} \frac{y}{\delta} e^{-y^2/2t} dy$$

$$= \frac{2}{\delta \sqrt{2\pi t}} \left[ -te^{-y^2/2t} \right]_{\delta}^{\infty} = \frac{2}{\delta \sqrt{2\pi}} \sqrt{te^{-\delta^2/2t}} \rightarrow as \ t \downarrow 0. \quad (3.31)$$

Therefore

$$\lim_{t \downarrow 0} |(p_t * g)(x) - g(x)| \leq \varepsilon \forall \varepsilon > 0.$$

Hence

$$\lim_{t \downarrow 0} |(p_t * g)(x) - g(x)| = 0.$$

Q.E.D.

**Lemma 3.3** If $f$ and $g$ are in $S(R)$ then

$$(f * g)(x) = \frac{1}{(2\pi)^n} \int f(\xi)e^{-i\xi \cdot x}g(\xi).$$
Proof:

\[(\hat{f} \ast g)(x) = \int_R \hat{f}(x - y)g(y)dy\]  \hspace{1cm} (3.32)
\[= \frac{1}{(2\pi)^n} \int_R \int_R f(\xi) e^{-i\xi \cdot (x-y)}g(y)d\xi dy\]  \hspace{1cm} (3.33)
\[= \frac{1}{(2\pi)^n} \int_R \int_R f(\xi) e^{-i\xi \cdot x} e^{i\xi \cdot y}g(y)dyd\xi\]  \hspace{1cm} (3.34)
\[= \frac{1}{(2\pi)^n} \int_R f(\xi) e^{-i\xi \cdot x} \hat{g}(\xi)d\xi.\]  \hspace{1cm} (3.35)

Q.E.D.

**Theorem 3.4** If \(g\) is in \(S(R)\) then

\[(\hat{g})(x) = g(x).\]

**Proof.** In the preceding lemma put \(f(\xi) = e^{-t\xi^2/2}\). Then, by (3.16), \(\hat{f}(x) = p_t(x)\). Thus

\[(p_t \ast g)(x) = \frac{1}{2\pi} \int_R e^{-t\xi^2} e^{-i\xi \cdot x} \hat{g}(\xi)d\xi.\]

Let \(t \downarrow 0\). Use Lemma B.2 on the left and Dom. Conv. theorem on the right to get

\[g(x) = \frac{1}{2\pi} \int_R e^{-i\xi \cdot x} \hat{g}(\xi)d\xi.\]

The asymmetric way in which the factor \(2\pi\) occurs in the inversion formula is sometimes a confusing nuisance. It is useful to distribute this factor among the forward and backward transforms. Define

\[(\mathcal{F}g)(\xi) = (2\pi)^{-1/2} \hat{g}(\xi)\]  \hspace{1cm} (3.36)

The factor in front of \(\hat{g}\) has clearly no effect on the one-to-one or ontoness property of the map \(g \to \hat{g}\). That is, \(\mathcal{F}\) is a one-to-one map of \(S(R)\) onto \(S(R)\). However the factor makes for a nice identity:

**Corollary 3.5** (*Plancherel formula for \(S\)*) If \(f\) and \(g\) are in \(S(R)\) then

\[(\mathcal{F}f, \mathcal{F}g) = (f, g)\]  \hspace{1cm} (3.37)
Explicitly, this says

\[ \int_R (\mathcal{F} f)(\xi)(\mathcal{F} g)(\xi) d\xi = \int_R f(x)g(x) dx \quad (3.38) \]

**Proof.** Phrasing this identity directly in terms of \( \hat{f} \) and \( \hat{g} \), it asserts that

\[ \int_R \hat{f}(\xi)\hat{g}(\xi) d\xi = (2\pi) \int_R f(x)g(x) dx \]

But \( \hat{g}(\xi) = \int g(x)e^{ix\cdot\xi} dx = \int g(x)e^{-ix\cdot\xi} dx \). Hence

\[ \int_R \hat{f}(\xi)\overline{\hat{g}(\xi)} d\xi = \int_R \int_R \hat{f}(\xi)\overline{g(x)}e^{-ix\cdot\xi} dxd\xi \quad (3.39) \]
\[ = \int_R \left( \int_R \hat{f}(\xi)\overline{g(x)}e^{-ix\cdot\xi} d\xi \right) dx \quad (3.40) \]
\[ = \int_R (2\pi) f(x)g(x) dx. \quad (3.41) \]

\[ \square \]

Q.E.D.

### 3.2 Extension of the Fourier transform to \( L^2 \) and other function spaces.

**Lemma 3.6** Let \( V \) be a dense linear subspace of a Banach space \( X \). Suppose that \( S : V \rightarrow Y \) is a bounded linear operator from \( V \) into a Banach space \( Y \). Then there exists a unique bounded linear operator \( T : X \rightarrow Y \) which extends \( S \).

**Proof.** Let \( M = \| S \|_{V \rightarrow Y} \). Suppose that \( x \in X \). Then there exists a sequence \( x_n \in V \) such that \( x_n \rightarrow x \). Hence \( \| Sx_n - Sx_k \|_Y \leq M \| x_n - x_k \|_V \rightarrow 0 \). Since \( Y \) is complete the Cauchy sequence \( Sx_n \) has a limit, \( y \). Define \( Tx = y \). Then

1. \( T \) is well defined. I.e. it is independent of the choice of Cauchy sequence. \( T \) and \( S \) agree on \( V \). (I leave these to you to show.)
2. \( T \) is linear on \( X \). (You can show this too.)
3. \( \| T \| \leq \| S \| \) and consequently \( \| T \| = \| S \| \). (Ditto).
4. The extension \( T \) is unique among all bounded linear operators extending \( S \). (Try your hand at this too.) \[ \square \]
Theorem 3.7 The Fourier transform $\mathcal{F} : S \to S$ has a unique continuous extension $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$. Moreover $\mathcal{F}$ is one-to-one, onto and isometric.

Proof. Put $g = f$ in (3.37) to deduce that $\|\mathcal{F}f\|_L^2 = \|f\|_L^2$ for all $f \in S$. Since $S$ is dense in $L^2(\mathbb{R})$ the lemma assures us that $\mathcal{F}$ has a unique continuous extension $\overline{\mathcal{F}}$ to $L^2$ and in fact $\|\overline{\mathcal{F}}\|_{L^2 \to L^2} = 1$. Furthermore, if $f_n \in S$ and $f_n \to f$ in $L^2$ norm then (3.37) and the proof of the lemma show that

$$\|\overline{\mathcal{F}}f\|_L^2 = \|f\|_L^2 \text{ for all } f \in L^2.$$  \hfill (3.42)

So $\overline{\mathcal{F}}$ is an isometry and is therefore automatically one-to-one. Now if $g_n \in L^2$ and $\overline{\mathcal{F}}g_n$ converges in $L^2$ to some element $h \in L^2$ then $\overline{\mathcal{F}}g_n$ is Cauchy and therefore, by isometry, $g_n$ is also Cauchy and hence has an $L^2$ limit $g$. But then $\overline{\mathcal{F}}g = \lim \overline{\mathcal{F}}g_n = h$. So the range of $\overline{\mathcal{F}}$ is a closed subspace of $L^2$. But the range of $\overline{\mathcal{F}}$ contains the range of $\mathcal{F}$, which is $S$ and which is itself dense in $L^2$. Hence the range of $\overline{\mathcal{F}}$ is all of $L^2$. \hfill \blacksquare

Terminology A linear map from one Hilbert space to another which is one-to-one, isometric, and surjective is called unitary. So the extended Fourier transform $\overline{\mathcal{F}}$ is a unitary operator from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$.

The prescription we’ve used to define $\overline{\mathcal{F}}$ on $L^2(\mathbb{R})$ has been rather indirect. Here is a more direct prescription.

Suppose that $f \in L^2(\mathbb{R})$ and is zero outside the interval $[-a,a]$. Choose a real number $b > a$. Then there exists a sequence $\phi_n \in C_c^\infty(\mathbb{R})$ such that all $\phi_n$ are zero outside $(-b,b)$ and converge to $f \in L^2((-b,b))$. Since $b$ is finite we know that $f$ is in $L^1((-b,b))$ and that the sequence converges to $f$ in $L^1((-b,b))$ also. Therefore, for each $\xi \in \mathbb{R}$

$$(2\pi)^{1/2}(\mathcal{F}\phi_n)(\xi) = \int_{-b}^{b} \phi_n(x)e^{ix\xi}dx$$

which converges, by the dominated convergence theorem, to $\int_{-b}^{b} f(x)e^{ix\xi}dx$, which, by the way is clearly a continuous function of $\xi$. But $\overline{\mathcal{F}}f$ is, by definition, the $L^2$ limit of the sequence $\mathcal{F}\phi_n$, which we have just seen converges pointwise to $\int_{-b}^{b} f(x)e^{ix\xi}dx$. Hence

$$(\mathcal{F}f)(\xi) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} f(x)e^{ix\xi}dx \hfill (3.43)$$
for almost all $\xi$. (Recall that the left side is just an $L^2$ limit.) Of course we may take (3.43) as the actual function representing the equivalence class on the left. It is customary to do so and we shall follow this custom.

Denote by $L^2_c(\mathbb{R})$ the subspace of $L^2(\mathbb{R})$ consisting of functions which are zero outside some bounded interval. We have seen that for these functions $\mathcal{F}f$ is just given explicitly by (3.43).

Next, let $g$ be an arbitrary function in $L^2(\mathbb{R})$. Define $g_n(x) = g(x)\chi_{(-n,n)}(x)$. Then (3.55) holds for $g_n$. But $g_n$ converges in $L^2(\mathbb{R})$ to $g$. (Use DCT.) Hence $\mathcal{F}g$ is the $L^2$ limit of $\mathcal{F}g_n$ because $\mathcal{F}$ is an isometry (and hence continuous.) Therefore we may write

$$\mathcal{F}_g(\xi) = \text{l.i.m.}_{n \to \infty} \int_{-n}^{n} g(x)e^{ix\xi}dx$$

(3.44)

where l.i.m. means “limit in the $L^2$ sense”, but is translated as “limit in mean”. (3.44) is the classical way (early 20th century) of describing the Fourier transform as a unitary operator on $L^2(\mathbb{R})$. You can’t get rid of that l.i.m. by just writing $\int_{-\infty}^{\infty} g(x)e^{ix\xi}dx$ because the integrand is not in $L^1(\mathbb{R})$ for a general function $g \in L^2(\mathbb{R})$. For example if $g(x) = (1 + |x|)^{-1}$ then $g \in L^2$ but not in $L^1$. Hence $\int_{-\infty}^{\infty} g(x)e^{ix\xi}dx$ is $+\infty$ at $\xi = 0$ and yet for a.e. $\xi$ the integrals in (3.44) converge as $n \to \infty$. (Of course you may need to drop down to a subsequence of $n$.) This results from cancellation in the oscillating integrand. Is that amazing, or what?

We now have specific formulas for $\mathcal{F}f$ for various spaces of functions, as follows.

$\mathcal{S}$, $L^2(\mathbb{R})$, $L^2(\mathbb{R})$, $L^1 \cap L^\infty$ (which is contained in $L^2$), and

$\{f \in L^1 \cap L^\infty \cap C(\mathbb{R}) : \hat{f} \in L^1(\mathbb{R})\}$

Given all the information that we have you can now verify that the inversion formula holds on the last space. Namely

$$f(x) = (2\pi)^{-1} \int_{\mathbb{R}} \hat{f}(\xi)e^{-ix\xi}d\xi \quad \forall \ x \in \mathbb{R}$$

(3.45)

if $f \in L^1 \cap L^\infty \cap C(\mathbb{R})$ and $\hat{f} \in L^1$.

Exercise: Prove (3.45).
3.3 The Fourier transform over $R^n$

The definitions, key formulas, theorems and proofs for the Fourier transform over $R^n$ are nearly identical to those over $R$. In this section we are going to summarize the results we’ve obtained so far and formulate them over $R^n$.

The Fourier transform of a complex valued function on $R^n$ is

$$\hat{f}(\xi) = \int_{R^n} e^{i\xi \cdot x} f(x) dx$$ (3.46) \text{F51}

Here $\xi$ runs over $R^n$. Just as in one dimension, one must pay some attention to the meaningfulness of this integral. But the ideas are similar.

As in one dimension, the Fourier transform over $R^n$ interchanges multiplication and differentiation. The analog of (F3.2) is

$$\frac{\partial}{\partial \xi_j} \hat{f}(\xi) = \int_{R^n} e^{i\xi \cdot x} \{ix_j f(x)\} dx.$$ (3.47) \text{F52}

So

$$(\text{Fourier transform of } \{ix_j f(x)\})(\xi) = \frac{\partial}{\partial \xi_j} \hat{f}(\xi).$$ (3.48) \text{F53}

An integration by parts (never mind the boundary terms) clearly gives the analog of (F3.4):

$$-i\xi_j \hat{f}(\xi) = \int_{R^n} e^{i\xi \cdot x} (\partial f / \partial x_j)(x) dx.$$ (3.49) \text{F54}

So

$$\partial f / \partial x_j(\xi) = -i\xi_j \hat{f}(\xi)$$ (3.50) \text{F55}

We will see later that these formulas allow one to solve some partial differential equations. Moreover in quantum mechanics these two formulas amount to the statement that the Fourier transform interchanges $P_j$ and $Q_j$ (momentum and position operators.)

We say that a function $f$ is in $L^1(R^n)$ if

$$\int_{R^n} |f(x)| d^n x < \infty$$ (3.51) \text{F56}

The formula (F3.46) makes perfectly good sense if $f \in L^1(R^n)$. But in the end we are going to give meaning to (F3.46) for a much larger class of functions and generalized functions, including e.g. delta functions and their derivatives. To this end we need the n-dimensional analog of $S$. 

43
**Definition 3.8** A function $f$ on $\mathbb{R}^n$ is said to be rapidly decreasing if

$$|x|^k |f(x)| \leq M_k, \quad k = 0, 1, 2, \ldots \quad (3.52)$$

for some real numbers $M_k$. In words: $|x|^k f(x)$ is bounded on $\mathbb{R}^n$ for each $k$.

In Exercise 4 you will have the opportunity to show that the condition (3.52) is equivalent to the statement that for any polynomial $p(x_1, \ldots, x_n)$ in $n$ real variables, the product $p(x)f(x)$ is bounded.

**Lemma 3.9** Any rapidly decreasing function on $\mathbb{R}^n$ is in $L^1(\mathbb{R}^n)$.

**Definition 3.10** $\mathcal{S}(\mathbb{R}^n)$ is the set of functions $f$ in $C^\infty(\mathbb{R}^n)$ such that $f$ and each of its partial derivatives are rapidly decreasing.

**Theorem 3.11**

a. If $f$ is in $\mathcal{S}(\mathbb{R}^n)$ then $\hat{f}$ is also in $\mathcal{S}(\mathbb{R}^n)$.

b. The map $f \to \hat{f}$ is a one-to-one linear map of $\mathcal{S}(\mathbb{R}^n)$ onto $\mathcal{S}(\mathbb{R}^n)$.

c. The inverse is given by the inversion formula

$$f(y) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-i\xi \cdot y} \hat{f}(\xi) d^n \xi. \quad (3.53)$$

**Definition 3.12** The convolution of two functions $f$ and $g$ on $\mathbb{R}^n$ is given by

$$(f \ast g)(x) = \int_{\mathbb{R}^n} f(x - y)g(y)d^n y \quad (3.54)$$

The important identity in the next theorem is the basis for the application of the Fourier transform to solution of partial differential equations.

**Theorem 3.13**

$$(\hat{f \ast g})(\xi) = \hat{f}(\xi)\hat{g}(\xi) \quad (3.55)$$

**Proof.** The proof consists of the following straight forward computation. A reader who is concerned about the validity of any of these steps should
simply assume that $f$ and $g$ are in $S(\mathbb{R}^n)$, although the identity is valid quite a bit more generally.

\[
\hat{f} \ast g(\xi) = \int_{\mathbb{R}^n} (f \ast g)(x) e^{ix \cdot \xi} dx = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y)g(x-y)dy e^{ix \cdot \xi} dx = \int_{\mathbb{R}^n} \hat{f}(\xi)g(x) e^{iy \cdot \xi} dy = \hat{f}(\xi)\hat{g}(\xi)
\]

\[\textbf{Theorem 3.14} \quad \text{(Plancherel formula)} \quad \text{Define } \mathcal{F}_{\phi} = (2\pi)^{-n/2} \hat{\phi}. \text{ Then } \left\| \mathcal{F}_{\phi} \right\|_{L^2(\mathbb{R}^n)} = \left\| \phi \right\|_{L^2(\mathbb{R}^n)} \]

This is the Plancherel formula. As a consequence $\mathcal{F}$ is a unitary operator on $L^2(\mathbb{R}^n)$.

This is a minor restatement of Corollary \text{corFP3.5} and extension to $\mathbb{R}^n$.

Finally, here are the $n$-dimensional analogs of the important Gaussian identities (3.14) - (3.16).

\textbf{Gaussian identities over } $\mathbb{R}^n$.

Let

\[
p_t(x) = \frac{1}{\sqrt{(2\pi t)^n}} e^{-|x|^2/2t} \quad x \in \mathbb{R}^n.
\]

Then

\[
\int_{\mathbb{R}^n} p_t(x) dx = 1
\]

\[
\hat{p}_t(\xi) = e^{-t|\xi|^2/2}
\]

\[
(e^{-\frac{t}{2}|\xi|^2})(x) = \frac{1}{\sqrt{(2\pi t)^n}} e^{-|x|^2/2t} = p_t(x).
\]
3.4 Tempered Distributions

Definition 3.15 A tempered distribution on $\mathbb{R}^n$ is a linear functional on $\mathcal{S}(\mathbb{R}^n)$.

Example 3.16 ($n = 1$) Suppose that $f : \mathbb{R} \to \mathbb{C}$ satisfies

$$|f(x)| \leq \text{const.}(1 + |x|^n)$$

for some $n \geq 0$ \hfill (3.66)

In this case we say that $f$ has polynomial growth. E.g. $x^3 \sin x$ has polynomial growth. (Take $n = 3$ in (3.66). Of course $n = 4$ will also do.) If $f$ has polynomial growth then the integral

$$L_f(\phi) \equiv \int_{\mathbb{R}} f(x) \phi(x) dx$$

(3.67) \hfill (F31)

exists when $\phi \in \mathcal{S}$ because $|f(x)\phi(x)| \leq \text{const.}(1 + |x|^n)(1 + |x|^{n+2})^{-1}$, which goes to zero at $\infty$ like $x^{-2}$ and so is integrable. Clearly $L_f$ is linear. Thus any function of polynomial growth determines in this natural way a linear functional on $\mathcal{S}$. Remember that we did not need polynomial growth of $f$ when we made $L_f$ into a linear functional on $\mathcal{D}$. So we have a “smaller” dual space now. But the delta distribution and its derivatives are in this smaller dual space anyway because

$$L_\delta(\phi) \equiv \delta(0)$$

is a meaningful linear functional on $\mathcal{S}$. (And similarly for $\delta^{(k)}$.)

Example 3.17 We can’t allow the function $f(x) = e^{2x^2}$ in (3.61) because the function $\phi(x) = e^{-x^2}$ is in $\mathcal{S}$ and so (3.67) wouldn’t make sense.

Definition 3.18 The Fourier transform of an element $L \in \mathcal{S}^\ast$ is defined by

$$\hat{L}(\phi) = L(\hat{\phi}) \text{ for } \phi \in \mathcal{S}.$$ \hfill (3.68) \hfill (F33)

Now you can see the virtue of using $\mathcal{S}$ as our new test function space: the right hand side of (3.68) makes sense precisely because $\hat{\phi}$ is back in $\mathcal{S}$ when $\phi$ is in $\mathcal{S}$. After all, $L$ is only defined on $\mathcal{S}$. Had we attempted to use $\mathcal{D}$ this definition would have failed because $\hat{\phi}$ is never in $\mathcal{D}$ when $\phi \in \mathcal{D}$ (if $\phi \neq 0$.)
Example 3.19 $\hat{L}_\delta = L_1$. (In the outside world this is usually written $\hat{\delta} = 1$.)

Proof. Let $\phi \in \mathcal{S}$. Then

\[
\begin{align*}
\hat{L}_\delta(\phi) &= L_\delta(\hat{\phi}) \text{ by def. } (3.68) \\
&= \hat{\phi}(0) \text{ by the def. of } L_\delta \quad (3.70) \\
&= \int_\mathbb{R} \phi(x) dx \text{ by the def. of Fourier trans. at 0 } (3.71) \\
&= L_1(\phi) \text{ by yet another definition } (3.72)
\end{align*}
\]

One thing to take away from this proof is that every step is just the application of some definition. There is no mysterious computation. Let this be a lesson to all of us!

Example 3.20

\[
\hat{L}_1 = 2\pi L_\delta \quad (3.73)
\]

After you have achieved a suitable state of sophistication you can write this as

\[
\hat{1} = 2\pi \delta.
\]

Or even as!!

\[
\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy} dx = \delta(y)
\]

(if you don’t feel too uncomfortable with the seeming nonsense of the left side. But you better get used to it. That’s how (3.73) is written in the outside world.)

Proof. of (3.73):

\[
\begin{align*}
\hat{L}_1(\phi) &= L_1(\hat{\phi}) \text{ by def. } (3.68) \\
&= \int_\mathbb{R} \hat{\phi}(\xi) d\xi \text{ by def. of } L_1 \quad (3.75) \\
&= 2\pi \phi(0) \text{ by the Fourier inversion formula } (3.76)
\end{align*}
\]

Notice that this time the last step uses something deep. ■
Consistency of the two definitions of the Fourier transform. If \( f \in L^1 \) then we already have a meaning for \( \hat{f} \). So is

\[
\hat{L_f} = L_f?
\]

Yes. Here is a (definition chasing) proof.

\[
\hat{L_f}(\phi) = L_f(\hat{\phi}) \tag{3.77}
\]

\[
= \int_{\mathbb{R}} f(\xi) \hat{\phi}(\xi) d\xi \tag{3.78}
\]

\[
= \int_{\mathbb{R}} f(\xi) \int_{\mathbb{R}} e^{ix\xi} \phi(x) dx d\xi \tag{3.79}
\]

\[
= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} f(\xi) e^{ix\xi} d\xi \right) \phi(x) dx \tag{3.80}
\]

\[
= \int_{\mathbb{R}} \hat{f}(x) \phi(x) dx \tag{3.81}
\]

\[
= L_f(\hat{\phi}) \tag{3.82}
\]

QED
3.5 Problems on the Fourier Transform

1. Find the Fourier transforms of the following functions:
   a) $e^{-|x|}$
   b) $e^{-x^2/2}$
   c) $xe^{-x^2/2}$
   d) $x^2e^{-x^2/2}$
   e) $f(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$
   f) $\frac{1}{1+x^2}$

2. For $\varphi$ in $\mathcal{S}(\mathbb{R}^2)$ define
   $$T(\varphi) = \int_{-\infty}^{\infty} x\varphi(x,0)\,dx.$$ (Note that this is an integral over a line — not over $\mathbb{R}^2$.)
   a) Show that $T$ is in $\mathcal{S}'(\mathbb{R}^2)$.
   b) Find the Fourier transform, $\hat{T}$, of $T$ explicitly (explicitly enough to do part c) without finding $\hat{\psi}$).
   c) Evaluate $\hat{T}(\psi)$ where
   $$\psi(\xi,\eta) = \frac{\xi e^{-(\xi^2+\eta^2)}}{1+\xi^2}$$
   d) Let $L = a\frac{\partial^2}{\partial \xi^2} + b\frac{\partial^2}{\partial \xi \partial \eta} + c\frac{\partial^2}{\partial \eta^2}$. Find all values of the real parameters $a$, $b$, $c$ such that $L\hat{T} = 0$.

3. Compute
   a. $\hat{\delta}'$
   b. $\hat{\delta}''$
   c. $\hat{L}_x$
4. Show that a function $f$ on $\mathbb{R}^n$ is rapidly decreasing if and only if, for every polynomial $p(x_1, \ldots, x_n)$, the function $p(x)f(x)$ is bounded.

5. Derive the identities $\mathbb{F}_{81}^{83}$ from the one-dimensional identities $\mathbb{F}_{11}^{13}$.