Teoretisk Fysik, KTH
September 5, 2006

5A1305 Fysikens matematiska metoder, ht06.

Introduction to Hilbert spaces. I.

My aim here is to give a down-to-earth introduction to the theory of Hilbert spaces which we will later discuss in more generality (and on a more abstract mathematical level). My aim is to stress the important mathematical ideas in a somewhat informal way.

WARNING: I typed this in very quickly, and I thus expect there are quite a few typos in the text: please let me know if you find some.

Fourier series: In Section 3.1.2 of our course book[1] (referred to as [KS] in the following) there is a short summary of the theory of Fourier series.
In the following I recall one important result from this, and I will discuss a useful way of looking at this result.
I recall that all continuous functions $f(x)$ defined on an interval of length L, $0 < x < L$ can be expanded in a sinus series as follows:

$$f(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x/L) \tag{1}$$

with coefficients given by the following integrals:

$$c_n = \frac{2}{L}\int_0^L \sin(n\pi x/L) f(x) dx. \tag{2}$$

(The theory of Fourier series allows for even functions $f(x)$ which have jumps and specifies in which sense the Fourier series represents this function, but we will ignore these details right now).
It is useful to give a name to the set of functions to which the result in Eqs. (1) and (2) applies: we call $\mathcal{D}$ the set of all continuous, real-valued functions f defined on the interval $(0, L)$ (the usual name for this set in mathematics is $C^1((0, L))$, but I use this shorter name for now). Note that we distinguish here f (as an element in this set $\mathcal{D}$) from $f(x)$ (which is the value of the function at some particular point x). As you know, this set is an important example of a vector space: given two functions f, g in $\mathcal{D}$ and two real constants a, b we can define $h = af + bg$ by

$$h(x) = af(x) + bg(x),$$

and this defines a continuous function, i.e. an element in $\mathcal{D}$, so that the axioms of a vector space are fulfilled (see [KS], Kap. H.1).

[1] [KS] G. Sparr, A Sparr: "Kontinuerliga system", Studentliterature, Lund (2000)

We now introduce the notation

$$u_n(x) = \sin(n\pi x/L), \tag{3}$$

which allows us to write (1) as

$$f = \sum_{n=1} c_n u_n. \tag{4}$$

This we can interpret as follows: the functions u_n, $n = 1, 2, \ldots$, are a set of special functions in $\mathcal{D}$, and the theory of Fourier series tell us that we can write every element in $\mathcal{D}$ as a linear superposition of these special functions.

Analogy with $\mathbb{R}^N$: The above is analog to writing vectors $\mathbf{v} = (v_1, v_2, \ldots, v_N)$ in $\mathbb{R}^N$ ($N = 2, 3, \ldots$) as a linear superposition of the following special vectors

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \quad \mathbf{e}_N = (0, 0, 0, \ldots, 1)$$

(often called *standard basis*), i.e.,

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_N\mathbf{e}_N = \sum_{n=1}^{N} v_n \mathbf{e}_n. \tag{5}$$

This analogy goes further: in the example of $\mathbb{R}^N$ we can compute the so-called *components* v_n of a vector $\mathbf{v}$ in this basis using the scalar product defined as usual,

$$(\mathbf{v}, \mathbf{w}) = \sum_{n=1}^{N} v_n w_n$$

for all vectors $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^N$. Recall that this scalar product satisfies certain rules (for example, $(\mathbf{v}, \mathbf{v})$ is always positive unless $\mathbf{v}$ is the zero vector, linearity etc.; see [KS], Kap. H.2). This formula for the components is as follows:

$$v_n = (\mathbf{e}_n, \mathbf{v}).$$

You know and understand this formula, of course. However, let me recall a simple mathematical proof of if (by "proof" I simply mean a chain of simple arguments to convince yourself or somebody else that something is true): An important property of the vectors $\mathbf{e}_n$ above is that they are *orthogonal*, i.e., the scalar product of two distinct such vectors is always zero:

$$(\mathbf{e}_n, \mathbf{e}_m) = 0 \qquad \text{for all } n \neq m$$

(this you can easily check). Now take the scalar product of both sides of Eq. (5) with some vector $\mathbf{e}_m$, $m = 1, 2, \ldots$:

$$(\mathbf{e}_m, \mathbf{v}) = (\mathbf{e}_m, \sum_n c_n \mathbf{e}_n) = \sum_n c_n (\mathbf{e}_m, \mathbf{e}_n)$$

where we used linearity of the scalar product. Due to orthogonality, all terms in the sum on the r.h.s. are zero except for the one where $n = m$, and we thus get

$$(\mathbf{e}_m, \mathbf{v}) = c_m (\mathbf{e}_m, \mathbf{e}_m),$$

or equivalently,

$$c_n = \frac{(\mathbf{e}_n, \mathbf{v})}{(\mathbf{e}_n, \mathbf{e}_n)} \tag{6}$$

for all $n = 1, 2, \ldots, N$ (I renamed the dummy index m to n). Since $(\mathbf{e}_m, \mathbf{e}_m) = 1$ we thus obtain the formula we wanted to prove.
The nice thing about this proof is that it immediately can be generalized: Assume you have *any* set of "special" vectors $\mathbf{f}_n$, $n = 1, 2, \ldots$, in $\mathbb{R}^N$ which are orthogonal: $(\mathbf{f}_n, \mathbf{f}_m) = 0$ for all $n \neq m$. Then the very same argument tells us that,

$$\mathbf{v} = \sum_{n=1}^{N} b_n \mathbf{f}_n \quad \text{with} \quad b_n = \frac{(\mathbf{f}_m, \mathbf{v})}{(\mathbf{f}_m, \mathbf{f}_m)}.$$

There are very many examples of such vectors, e.g. for $N = 2$

$$\mathbf{f}_1 = (a\cos(c), a\sin(c)), \quad \mathbf{f}_2 = (-b\sin(c), b\cos(c))$$

for arbitrary non-zero constants a, b and real c.

Orthogonal basis: I now claim: Formula (2) above is perfectly analogous to (6), and its proof is essentially the same. Indeed, we can define the following "scalar product" of two functions f and g in our function space $\mathcal{D}$:

$$(f, g) = \int_0^L f(x)g(x)dx, \tag{7}$$

and with that notation we can write (2) short as follows

$$c_n = \frac{(u_n, f)}{(u_n, u_n)} \tag{8}$$

(convince yourself that this is true, recalling that $\int_0^L \sin^2(n\pi x/L)dx = L/2$). This is no coincidence: the "scalar product" in $\mathcal{D}$ obeys all axioms of a scalar product in a generalized sense (and we thus will drop in the following the quotation marks), and our special functions u_n above are orthogonal in the very same sense as above:

$$(u_n, u_m) = \int_0^L \sin(n\pi x/L)\sin(m\pi x/L)dx = 0 \quad \forall n \neq m$$

(check that!). Thus we can derive (8) from (4) by the very same arguments used to derive (6) from (5).
There is an important limitation in this argument: we obtain (8) *assuming* (4), but is it true that arbitrary functions in $\mathcal{D}$ have a representation as in (4)? The theory of Fourier series tells us "yes", but this is a very non-trivial statement.
To appreciate this let's again consider the example $\mathbb{R}^N$: In this case we could also choose a system or orthogonal vectors $\mathbf{f}_n$ where $n = 1, 2, \ldots, M$ with $M < N$ (one can show that $M > N$ is not possible). In this case the vector

$$\mathbf{v}' = \sum_{n=1}^{M} b_n \mathbf{f}_n \quad \text{with} \quad b_n = \frac{(\mathbf{f}_m, \mathbf{v})}{(\mathbf{f}_m, \mathbf{f}_m)} \tag{9}$$

is in general not identical with $\mathbf{v}$. What is the meaning of this vector? The answer is: $\mathbf{v}'$ *is the best possible approximation of* $\mathbf{v}$ *by a linear combination of the vectors* $\mathbf{f}_n$, $n = 1, 2, \ldots, M$, *in the sense that the "error"*

$$(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w}), \quad \mathbf{w} = \sum_{n=1}^{M} b_n \mathbf{f}_n$$

as a function of the parameters b_n *acquires its minimum for the* b_n *given in (9)* (this is easy to prove: using orthogonal and the properties of the scalar product we find that

$$(\mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w}) = (\mathbf{v}, \mathbf{v}) - 2 \sum_{n=1}^{M} b_n (\mathbf{f}_n, \mathbf{v}) + \sum_{n}^{M} b_n^2 (\mathbf{f}_n, \mathbf{f}_n)$$

and it is easy to compute the minimum of the function of the b_n on the r.h.s.). In the important special case where $M = N$ this approximation becomes perfect, i.e., $\mathbf{v}' = \mathbf{v}$. One says that an orthogonal system of functions $\mathbf{f}_n$, $n = 1, 2, \ldots, M$, is *incomplete* for $M < N$ and *complete* for $M = N$: This is what is meant by saying that $\mathbb{R}^N$ has the dimension N.
In our function space $\mathcal{D}$ we have an infinite number of orthogonal functions u_n: $\mathcal{D}$ is an example of an infinite dimensional vector space. The question of completeness is therefore much more subtle (note that having an infinite number of orthonormal functions does not imply that this system is complete: if we start from a complete system and throw away half of the function we still have an infinite number of functions but the system no longer is complete).
However, the discussion above can be taken over word by word to show that orthogonality of a system of function v_n, $n = 1, 2, \ldots$, implies that the function

$$f' = \sum_{n=1}^{M} b_n v_n \;\text{ with }\; b_n = \frac{(v_n, f)}{(v_n, v_n)}$$

is the best possible approximation of an arbitrary function f in $\mathcal{D}$ by a finite linear combination of the functions v_n, $n = 1, 2, \ldots$, and it is natural to *define* that any such system is complete if the "error"

$$(f - f', f - f')$$

converges to zero as $M \to \infty$. The non-trivial statement in the formulas (1) and (2) thus is the statement about completeness of the orthonormal system of functions u_n, $n = 1, 2, \ldots$, in $\mathcal{D}$. Such a complete orthonormal system of functions is called a *basis*. We will later discuss many more important examples of such a basis, but at this point let me only mention the functions

$$v_n(x) = \cos(n\pi x/L), \quad n = 0, 1, 2, \ldots$$

which are also orthonormal and complete in $\mathcal{D}$: convince yourself that the formula

$$f = \sum_{n=1}^{\infty} b_n v_n, \quad b_n = \frac{(v_n, f)}{(v_n, v_n)}$$

is equivalent to the well-known expansion of functions in $\mathcal{D}$ in a cosine series.

Comments: The *analogy* between $\mathcal{D}$ and $\mathbb{R}^N$ above is no coincidence but a consequence of the fact that there exists a general theory with $\mathcal{D}$ and $\mathbb{R}^N$ as special cases: the theory of Hilbert spaces. This general theory has many more important special cases which we will discuss in this course. Our course book has a concise summary of this general theory ([KS], Kap. H), and above I tried to show you a useful strategy to learn this theory: if you have a general result, consider a simple example which you know well (e.g. the case $\mathbb{R}^N$) and look what the result means in that special case. If you understand well the special case you made a big step towards understanding the general case.
Above I discussed the important notion of an orthogonal basis in a Hilbert space. The next important question is how to obtain such a basis. Again, the answer to this question is analog to the answer in $\mathbb{R}^N$: In the latter case we can get a basis from a *symmetric* $N \times N$ *matrix* $A = (A_{mn})_{m,n=1}^N$ (i.e., the matrix elements A_{nm} all are real and $A_{nm} = A_{mn}$). We know that such a matrix always has N orthogonal eigenvectors $\mathbf{f}_n$, and the corresponding eigenvalues λ_n all are real: they solve the eigenvalue equation

$$A\mathbf{f} = \lambda \mathbf{f}$$

where $(A\mathbf{v})_n = \sum_{m=1}^N A_{nm} v_n$ (spectral theorem for symmetric matrices).
The corresponding object in $\mathcal{D}$ is a symmetric operator, and important examples of these are given by *Sturm-Liouville operators*. For example, the Sturm-Liuville problem: "*find all the functions f in $\mathcal{D}$ satisfying the conditions*

$$f''(x) + \lambda f(x) = 0, \quad f(0) = f(L)$$

for some constant λ" is analog to the problem "*find the eigenvectors and eigenvalues of the symmetric matrix A*". Indeed, it is natural to regard the above as the eigenvalue equation of the differential operator

$$\mathcal{L} = -\frac{d^2}{dx^2}$$

where the domain of definition of $\mathcal{L}$ is not all of $\mathcal{D}$ but only the subset of function which can be differentiated twice and which valish in the endpoints $x = 0, L$ (it turns out that one needs to impose such restrictin to specify such a *symmetric operator* which has properties analog to the ones of a symmetric matrix). The Sturm-Liouville problem is simple enough that we can solve it by explicit computations:

$$f(x) = \sin(n\pi x/L), \quad \lambda = (n\pi/L)^2, \quad n = 1, 2, \ldots,$$

i.e. the eigenfunctions of $\mathcal{L}$ are exactly our functions u_n!
Again, this is no coincidence: according to the Sturm-Liouville theorem, the eigenfunctions of any Sturm-Liouville operator comprise a basis.
I plan to give a more detailed informal discussion of this some other time.