Chapter 6

Mathematical Foundations

6.1 Introduction

In this chapter we present a mathematical interlude to provide a more formal language which will allow us to set the mathematical foundations of quantum mechanics. Sections 6.2 and 6.3 contain a discussion of Hilbert space and should be read by all those not familiar with this concept. The remaining sections are somewhat more mathematical in nature and are provided for those who desire more rigour.

In section 6.4 we discuss linear operators in Hilbert space and introduce the concept of self-adjointness. The Cayley transform is introduced in section 6.5 and used to classify all self-adjoint extensions of a symmetric operator. Section 6.6 is devoted to some examples illustrating the results of section 6.5. More examples are also provided in section 6.7.

6.2 Geometry of Hilbert Space

The language we have used so far is one of wave-functions or state vectors and operators on these wave-functions. There is a ready-made mathematical language for this. This is the language of Hilbert space and quantum mechanics is naturally formulated in Hilbert space. Actually, for practical purposes Hilbert space is too small and the appropriate generalization is to a so-called rigged Hilbert space. We shall ignore this for the time being and discuss it briefly in Chapter 8, where we shall also give references for those who are interested in more of the details. What is Hilbert space? First of all it is a vector space analogous to the usual Euclidean spaces $\mathcal{E}_3$ or $\mathcal{E}_n$, however unlike $\mathcal{E}_3$ which is 3-dimensional and $\mathcal{E}_n$ which is n-dimensional, Hilbert space is ∞-dimensional. As a conceptual model of Hilbert space $\mathcal{H}$ one can think of taking some representation of a vector in $\mathcal{E}_n$, say $(a_1, \ldots, a_n)$ and writing it $(a_1, \ldots, a_n, 0, 0 \ldots, 0)$. Then, letting n increase without limit we arrive at the notion of $\mathcal{H}$. However, unlike all ordered n-tuples which automatically can be considered in $\mathcal{E}_n$, not all

infinite ordered sequences can be considered to belong to $\mathcal{H}$. The reason for this is that not all of them have finite "length" and we wish to include in $\mathcal{H}$ only elements of finite length. This is tantamount to saying in the language of wavefunctions that we want the wavefunction to be square integrable. We start by listing those properties of $\mathcal{E}_n$ which remain true in the transition to $\mathcal{H}$. The elements of the space are vectors f on which certain operations are defined.

1) Scalar multiplication : If $f \in \mathcal{H}$ and λ is a complex number then $\lambda f \in \mathcal{H}$.

2) Addition: If f_1 and $f_2 \in \mathcal{H}$ then $f_1 + f_2 \in \mathcal{H}$. Thus, combining 1) and 2) we see that all finite linear combinations of elements in $\mathcal{H}$ belong to $\mathcal{H}$.

3) Inner Product: In $\mathcal{H}$ there is defined an inner product (f, g) for all $f, g \in \mathcal{H}$. This inner product maps elements of $\mathcal{H}$ into complex numbers and satisfies the following conditions.
a)

$$(f, g) = (g, f)^* \tag{6.2.1}$$

where the star means "complex conjugation".
b)

$$(\lambda_1 f, \lambda_2 g) = \lambda_1^* \lambda_2 (f, g) \tag{6.2.2}$$

where λ_1 and λ_2 are complex numbers. We sometimes also use a bar over a number to indicate complex conjugation.
c)

$$\begin{aligned} (f_1 + f_2, g) &= (f_1, g) + (f_2, g) \\ (f, g_1 + g_2) &= (f, g_1) + (f, g_2) \ . \end{aligned} \tag{6.2.3}$$

d)

$$|(f, g)|^2 \leq (f, f)(g, g) \ . \tag{6.2.4}$$

This last inequality is the Schwarz inequality. We shall later derive this.

In terms of the inner product we define the "length" or norm of a vector in the usual way by

$$\| f \|^2 = (f, f) \ . \tag{6.2.5}$$

It is also possible to define orthogonality using the inner product. Thus, f is orthogonal to g if and only if (abbreviated iff)

$$(f, g) = 0 \ . \tag{6.2.6}$$

A set of vectors $\{f_i\}$ is orthonormal that is orthogonal and normal iff

$$(f_i, f_j) = \delta_{ij} \ . \tag{6.2.7}$$

We now come to some of the differences. For this we need two definitions.

Definition: A set $\{f_i\}$ of vectors is *complete* iff any vector in $\mathcal{H}$ can be written as a linear combination of vectors from the set $\{f_i\}$. A complete set of orthonormal vectors forms a *basis*.

Example:
Consider the Euclidean space $\mathcal{E}_3$ and choose three orthonormal vectors $(\hat{e}_1, \hat{e}_2, \hat{e}_3)$. Thus, $(\hat{e}_i, \hat{e}_j) = \delta_{ij}$. Then any vector $f \in \mathcal{E}_3$ can be written

$$f = \sum_{i=1}^{3} \lambda_i \hat{e}_i \tag{6.2.8}$$

and in fact

$$\lambda_i = (\hat{e}_i, f) \, . \tag{6.2.9}$$

The $\hat{e}_i$ are obviously complete and form a basis in $\mathcal{E}_3$. On the other hand, if we choose just two of these vectors, say $\hat{e}_1$ and $\hat{e}_2$, they do not form a complete set since for example any vector with a component along $\hat{e}_3$ cannot be expressed in terms of just the first two.

The λ_i are usually called the components of the vector. If we then agree to keep the basis fixed, we can suppress the basis vectors and write $f = (\lambda_1, \lambda_2, \lambda_2)$. In this manner we establish a one-one correspondence between vectors in $\mathcal{E}_3$ and ordered triplets. The norm of f is given by

$$\begin{aligned} \| f \|^2 &= \sum_{i,j=1}^{3} \lambda_i^* \lambda_j (\hat{e}_i, \hat{e}_j) \\ &= \sum_{i,j=1}^{3} \lambda_i^* \lambda_j \delta_{ij} \\ &= \sum_{i=1}^{3} |\lambda_i|^2 \, . \end{aligned} \tag{6.2.10}$$

This formalism above is all exceedingly trivial and you may wonder why bother. The reason is to establish a precise formalism so that when the situation becomes complicated we can rely on the formalism and not just our intuition.

Another use for the word *complete* is in the description of a vector space. This concept again is trivial for $\mathcal{E}_n$ but is non-trivial for $\mathcal{H}$. Consider a sequence of vectors $f_1, f_2, f_3, \ldots$. Furthermore, suppose that for every $\epsilon > 0$ we can find an n such that for any finite m

$$\| f_{n+m} - f_n \| < \epsilon \, . \tag{6.2.11}$$

This is just a statement of the Cauchy criterion for convergence using the norm $\| \cdot \|$ rather than the absolute value as is the usual case for numerical sequences. We call such a sequence a *Cauchy sequence*. Now if $\mathcal{H}$ in this case is finite

dimensional say $\mathcal{E}_n$ then it is trivial to show that the limit of the sequence exists and is a vector in $\mathcal{E}_n$. This property that all Cauchy sequences have a limit in $\mathcal{E}_n$ is stated by saying that $\mathcal{E}_n$ is complete. In fact all finite dimensional vector spaces are complete. This is also true for ∞-dimensional Hilbert spaces. There is a deep theorem of analysis known as the *Riesz-Fischer Theorem* which states that the space used by us (called $\mathcal{L}_2$ by mathematicians) is complete and thus a Hilbert space.

6.3 $\mathcal{L}_2$: A Model Hilbert Space

The elements of $\mathcal{L}_2$ are square-integrable complex-valued functions $f(x)$ of a real variable x. More generally x is a vector in some real finite vector space so that f is a function of n real variables. This generalization has no effect whatever on the ensuing statements and so we ignore it. The norm $\| \cdot \|$ in $\mathcal{L}_2$ is defined by

$$\| f \|^2 = (f,f) = \int f^*(x) f(x)\, dx \,. \tag{6.3.12}$$

The range of integration in (6.3.12) is over the full range of the variable x. Thus, if x is unrestricted, the integral runs from $-\infty$ to ∞. It is trivial to check our first two conditions for elements in $\mathcal{L}_2$. Thus,

$$f \in \mathcal{L}_2 \;\Rightarrow\; \lambda f \in \mathcal{L}_2 \tag{6.3.13}$$

and

$$f_1 \,,\, f_2 \in \mathcal{L}_2 \;\Rightarrow\; f_1 + f_2 \in \mathcal{L}_2 \,. \tag{6.3.14}$$

Furthermore, as defined by (6.3.12) the inner product obviously satisfies conditions a), b) and c).

a)

$$(f,g) = \int f^* g\, dx = \left(\int f g^*\, dx \right)^* = (g,f)^* \tag{6.3.15}$$

b)

$$(\lambda_1 f, \lambda_2 g) = \int (\lambda_1 f)^* (\lambda_2 g)\, dx = \lambda_1^* \lambda_2 \int f^* g\, dx = \lambda_1^* \lambda_2 (f,g) \tag{6.3.16}$$

c)

$$\begin{aligned} (f_1 + f_2, g) &= \int (f_1 + f_2)^* g\, dx = \int f_1^* g\, dx + \int f_2^* g\, dx \\ &= (f_1, g) + (f_2, g) \end{aligned} \tag{6.3.17}$$

and

$$\begin{aligned} (f, g_1 + g_2) &= \int f^* (g_1 + g_2)\, dx = \int f^* g_1\, dx + \int f^* g_2\, dx \\ &= (f, g_1) + (f, g_2) \,. \end{aligned} \tag{6.3.18}$$

The only condition left to verify on the inner product is the Schwarz inequality. To do this consider the vector

$$h = f + \lambda(g, f)g \tag{6.3.19}$$

where $f, g \in \mathcal{L}_2$ and λ is a real number. Then,

$$(h, h) = \| h \|^2 \geq 0 \ . \tag{6.3.20}$$

Therefore,

$$\begin{aligned} 0 &\leq (f + \lambda(g, f)g, f + \lambda(g, f)g) \\ &= (f, f) + 2\lambda|(f, g)|^2 + \lambda^2|(f, g)|^2(g, g) \ . \end{aligned} \tag{6.3.21}$$

This means that the quadratic polynomial in λ cannot have two real distinct zeros and hence that the discriminant is negative, giving

$$|(f, g)|^4 - |(f, g)|^2(f, f)(g, g) \leq 0 \ . \tag{6.3.22}$$

The equality sign obviously applies when $(f, g) = 0$ or $f = \mu g$. Thus, even if $(f, g) = 0$, we get

$$|(f, g)|^2 \leq (f, f)(g, g) \tag{6.3.23}$$

as required.

Orthogonality is still given by: f is orthogonal to g iff

$$(f, g) = 0 \ . \tag{6.3.24}$$

An example of two orthogonal vectors in $\mathcal{L}_2$ is:

$$f = \frac{1}{\pi^{1/4}} e^{-x^2/2} \quad g = \frac{2x}{\pi^{1/4}} e^{-x^2/2} \tag{6.3.25}$$

where the range of intergration is $(-\infty, \infty)$. The functions displayed are the first two hermite functions. As we see later, it is a general fact that eigenfunctions corresponding to different eigenvalues are orthogonal. In fact for physical Hamiltonians, the eigenfunctions properly normalized can be taken as a basis. This is an important fact since it implies that these eigenvectors form a complete set. Although we do not prove the Riesz-Fischer Theorem we restate it here. Consider a sequence of functions $f_1, f_2, \ldots$ all of which belong to $\mathcal{L}_2$. Furthermore, let this be a Cauchy sequence. This means that given any $\epsilon > 0$ we can find an $n > 0$ such that

$$\left[\int |f_{n+m} - f_n|^2 \, dx\right]^{1/2} < \epsilon \ . \tag{6.3.26}$$

Then the Riesz-Fischer Theorem asserts that
a) $\lim_{m\to\infty} f_m(x) = f(x)$
exists and, most important,

b) $f(x) \in \mathcal{L}_2$.
That is,

$$\int |f(x)|^2 \, dx < \infty \ . \tag{6.3.27}$$

This guarantees that we can take limits of sequences in $\mathcal{L}_2$. Clearly $\mathcal{L}_2$ is the model for the quantum mechanical Hilbert spaces. All Hilbert spaces are complete by definition and so all limits of Cauchy sequences are again elements of the space. We shall see that the inner product plays an exceedingly important role in the physical interpretation of quantum mechanics. There is one more technical point, namely that for any set of linearly independent vectors $\{f_j\}$ it is possible to construct an orthonormal set of vectors $\{e_j\}$ which span the same space. The orthogonalization process is called the *Schmidt Orthogonalization Procedure*. The proof is by construction. Choose one of the f_j say f_1. Then,

$$e_1 = \frac{f_1}{\| f_1 \|} \ . \tag{6.3.28}$$

Now form

$$g_2 = f_2 - (e_1, f_2)e_1 \tag{6.3.29}$$

and

$$e_2 = \frac{g_2}{\| g_2 \|} \ . \tag{6.3.30}$$

Clearly, e_1 and e_2 are orthonormal. Next form

$$g_3 = f_3 - (e_1, f_3)e_1 - (e_2, f_3)e_2 \tag{6.3.31}$$

and

$$e_3 = \frac{g_3}{\| g_3 \|} \ . \tag{6.3.32}$$

The process is now obvious.

We next turn to another aspect of Hilbert space, namely operators.

6.4 Operators on Hilbert Space: Mainly Definitions

An operator on Hilbert space is a mapping which maps certain elements of $\mathcal{H}$ into $\mathcal{H}$. Thus, if A is an operator with *domain* $D_A \subset \mathcal{H}$ then for all $f \in D_A$

$$g = Af \in \mathcal{H} \ . \tag{6.4.33}$$

The domain D_A consists simply of all those vectors in $\mathcal{H}$ such that the result of operating with A on a vector in D_A is again a vector in $\mathcal{H}$. Thus, for example, if $\mathcal{H}$ is the space $\mathcal{L}_2$ and A is the operator x^2 (multiplication by x^2) then

$$(Af)(x) = x^2 f(x) \ . \tag{6.4.34}$$

Clearly even if $f \in \mathcal{L}_2$ not all functions $x^2 f(x)$ are in $\mathcal{L}_2$. For example

$$f(x) = (x^2 + a^2)^{-m} \tag{6.4.35}$$

is in $\mathcal{L}_2$ for $\Re(m) > 1/4$. But, $x^2 f(x) = x^2(x^2 + a^2)^{-m}$ is not in $\mathcal{L}_2$ unless $\Re(m) > 5/4$.

Again we shall only be interested in linear operators. Thus,

$$A(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 A f_1 + \lambda_2 A f_2 \, . \tag{6.4.36}$$

For example, the operator x^2 defined above is linear and so are the operators

$$pf = \frac{\hbar}{i} \frac{df}{dx} \tag{6.4.37}$$

and

$$(Kf)(x) = \int_{-\infty}^{\infty} K(x,y) f(y) \, dy \, . \tag{6.4.38}$$

On the other hand, $\log f$ and $\sqrt{f}$ are definitely not linear operators acting on the function f . Linear operators are also familiar in finite dimensional vector spaces. They are usually represented by matrices in this case. There is an analogous representation for operators in Hilbert space. Formulated in this way quantum mechanics is called matrix mechanics to distinguish it from the Schrödinger formulation or wave mechanics. Both formulations are just two different mathematical ways of looking at the same thing. We examine matrix mechanics after we have developed all the necessary mathematical machinery. To illustrate the matrix operator formalism we first derive the form of the most general linear operator on a finite vector space say $\mathcal{E}_n$. Let A be such an operator. Call

$$g = Af \tag{6.4.39}$$

and consider taking for f different elements of a basis set $\{e_i\}$. Thus, let

$$g_i = Ae_i \, . \tag{6.4.40}$$

Then writing

$$f = \sum_i \lambda_i e_i \tag{6.4.41}$$

and

$$g = \sum_i \mu_i e_i \tag{6.4.42}$$

we get by linearity:

$$\sum_i \mu_i e_i = \sum_i \lambda_i A e_i = \sum_i \lambda_i g_i \, . \tag{6.4.43}$$

Now using

$$(e_i, e_j) = \delta_{ij} \tag{6.4.44}$$

and taking inner products in (6.4.43) we get:

$$\mu_j = \sum_i \lambda_i (e_j, Ae_i) = \sum_i \lambda_i (e_j, g_i) \ . \tag{6.4.45}$$

This means that the operator A is completely determined in this basis by the matrix of numbers

$$A_{ji} = (e_j, Ae_i) \ . \tag{6.4.46}$$

Conversely if we are given a matrix $(n \times n)$ then it can always be used to define a linear operator according to (6.4.46). Thus, as stated previously, the most general linear operator on $\mathcal{E}_n$ can be considered to be an $(n \times n)$ matrix. With only some attention to details the same argument will go through for an ∞-dimensional Hilbert space. Just as a matrix algebra is possible, so an algebra of linear operators is generally possible. It is only necessary to pay due attention to such things as domains of the operators. Let A be an operator on $\mathcal{H}$ with domain D_A. Then λA is also an operator on $\mathcal{H}$ with domain D_A and acts as follows:

$$(\lambda A) f = \lambda (Af) \ . \tag{6.4.47}$$

This is almost too obvious. If A, B are operators with domains D_A and D_B respectively, then $A + B$ is an operator with domain $D_A \cap D_B$ defined by

$$(A + B) f = Af + Bf \ . \tag{6.4.48}$$

The *range* R_A of an operator A is defined as the set of all vectors obtained by operating with A on elements in D_A. Symbolically,

$$R_A = AD_A \ . \tag{6.4.49}$$

Then if A, B are operators with domains D_A and D_B and $R_A \subset D_B$ we can define the product operator BA according to

$$(BA) f = B(Af) \ . \tag{6.4.50}$$

This is well defined since by assumption

$$f \in D_A \ . \tag{6.4.51}$$

Therefore

$$Af \in R_A \subset D_B \tag{6.4.52}$$

and hence

$$Af \in D_B \ . \tag{6.4.53}$$

Conversely if $R_B \subset D_A$ we can define the product

$$(AB)f = A(Bf) \ . \tag{6.4.54}$$

This points out the interesting possibility that although BA may exist as an operator AB might not, and conversely. Another property which many operators of physical interest possess is hermiticity. Actually the interesting property is self-adjointness and we shall examine these two properties in some detail to bring out the difference. First we need some definitions. Let A be an operator on $\mathcal{H}$ and $f \in D_A$. Then consider the expression (g, Af). If for some $g \in \mathcal{H}$ we find that there is an $h \in \mathcal{H}$ such that

$$(g, Af) = (h, f) \tag{6.4.55}$$

for all $f \in D_A$ then we define the adjoint operator $A^\dagger$ of A by

$$h = A^\dagger g \tag{6.4.56}$$

with domain $D_{A^\dagger}$ = the set of all g for which (6.4.55) holds. In that case we can write

$$(g, Af) = (A^\dagger g, f) \ . \tag{6.4.57}$$

Note that the element h in (6.4.55) is defined uniquely by g if the domain D_A contains sufficiently many vectors. The precise statement of this is that D_A is *dense* in $\mathcal{H}$. For our purposes a set in $\mathcal{H}$ is dense if any element in $\mathcal{H}$ can be approximated arbitrarily closely by an element from this set. Thus, D_A is dense if for any $f \in \mathcal{H}$ there exists a $g \in D_A$ such that given

$$\epsilon > 0 \quad \| f - g \| < \epsilon \ . \tag{6.4.58}$$

In this case the proof that h is unique is trivial. For, assume there is another such vector h'. Then,

$$(g, Af) = (h', f) \tag{6.4.59}$$

as well. Combining this with (6.4.55) we get

$$(h - h', f) = 0 \ . \tag{6.4.60}$$

Thus, $h - h'$ is orthogonal to every vector in D_A. But D_A is dense in $\mathcal{H}$ so that for any vector $g \in \mathcal{H}$

$$|(h - h', g)| < \epsilon \ . \tag{6.4.61}$$

This is possible only if $h = h'$. The adjoint operator is also a linear operator as is immediately obvious. Now again let A be an operator in $\mathcal{H}$ with domain D_A then A is *hermitian* if for all $f, g \in D_A$

$$(Af, g) = (f, Ag) \ . \tag{6.4.62}$$

An operator A is *symmetric* if it is hermitian and its domain of definition D_A is dense in $\mathcal{H}$. From the definition of $A^\dagger$ it then follows that for a symmetric operator A

$$D_A \subset D_{A^\dagger} \tag{6.4.63}$$

as we show for an example. If in addition

$$D_A = D_{A^\dagger} \tag{6.4.64}$$

or as this implies

$$A = A^\dagger \ . \tag{6.4.65}$$

Then, A is *self-adjoint*. To make this less abstract consider the momentum operator

$$p = \frac{\hbar}{i}\frac{d}{dx} \tag{6.4.66}$$

defined on the Hilbert space $\mathcal{L}_2(a,b)$ of functions square-integrable on the interval $[a,b]$. As domain of this operator we choose

$$D_p = \left\{ f \in \mathcal{L}_2(a,b) \mid \frac{df}{dx} \text{ is bounded on } (a,b)\ ,\ f(a) = f(b) = 0 \right\} \ . \tag{6.4.67}$$

With this definition it is easy to see that p is hermitian and in fact symmetric. The domain D_p is dense in $\mathcal{L}_2(a,b)$. Thus, if p is hermitian, it is symmetric. To see hermiticity let

$$f \in D_p \quad , \quad g \in D_p \ . \tag{6.4.68}$$

Then,

$$\begin{aligned} (f, pg) &= \int_a^b f^*(x)\frac{\hbar}{i}\frac{dg(x)}{dx}\,dx \\ &= \frac{\hbar}{i} f^*(x)g(x)\Big|_a^b + \int_a^b \left(\frac{\hbar}{i}\frac{df(x)}{dx}\right)^* g(x)\,dx \ . \end{aligned} \tag{6.4.69}$$

Since $f^*(a) = f^*(b) = g(a) = g(b) = 0$, the term obtained from integration by parts vanishes and so we have

$$(f, pg) = (pf, g) \ . \tag{6.4.70}$$

Thus, p is hermitian (symmetric) as claimed. On the other hand $p \neq p^\dagger$ since, as we now show, the domain $D_{p^\dagger}$ of $p^\dagger$ is much larger than the domain D_p of p, i.e. $D_p \subset D_{p^\dagger}$, but $D_p \neq D_{p^\dagger}$. This means that D_p is a proper subset of $D_{p^\dagger}$. To see this consider any $g \in D_p$ and let f be any function whose derivative is bounded over (a,b) and such that

$$f(b) = e^{i\theta} f(a) \tag{6.4.71}$$

where θ is a constant. Then by a computation, identical to the one above, we again find

$$(f, pg) = (pf, g) \ . \tag{6.4.72}$$

Thus, as a differential operator

$$p^\dagger = \frac{\hbar}{i}\frac{d}{dx} \tag{6.4.73}$$

but the domain of $p^\dagger$ is larger than the domain of p. It is furthermore easy to check that if we define

$$D_{p^\dagger} = \left\{ f \in \mathcal{L}_2(a,b) \ \middle| \ \frac{df}{dx} \text{ is bounded on } (a,b) \text{ and } f(b) = e^{i\theta} f(a) \right\} \tag{6.4.74}$$

then $p^\dagger$ is also symmetric. Thus, we say that $p^\dagger$ is a symmetric extension of p. In fact one can check that $p^\dagger$ is self-adjoint because $D_{p^\dagger} = D_{p^{\dagger\dagger}}$. Thus, we have a self-adjoint extension of p.

A symmetric operator A is essentially self-adjoint if $A^{\dagger\dagger}$ is self-adjoint. What this means is that although A itself is not necessarily self-adjoint there is a unique way to extend it to a self-adjoint operator. That $A^{\dagger\dagger}$ is an extension of A follows from

$$D_A \subset D_{A^\dagger} \subset D_{A^{\dagger\dagger}} \ . \tag{6.4.75}$$

The operator p discussed above is not essentially self-adjoint because, for each value of the parameter θ used to define $D_{p^\dagger}$, we get a different self-adjoint extension. This means that the physical results we get for different values of θ are different. A neat way to say this is that different self-adjoint extensions give different physics. Rather than being a nuisance, this makes the structure of quantum mechanics much richer.

6.5 Cayley Transform: Self-Adjoint Operators

We now examine under what conditions a general symmetric operator possesses self-adjoint extensions and how many. To do this we need some more machinery. The operation analogous to a rotation in a Euclidean space $\mathcal{E}_n$ is a unitary transformation in $\mathcal{H}$. The characteristic property of a rotation in $\mathcal{E}_n$ is that it preserves length and angles or more succinctly, it preserves the inner product. This is also its characteristic in $\mathcal{H}$.

Definition U is unitary iff $D_U = R_U = \mathcal{H}$ and

$$(Uf, Uf) = (f, f) \ . \tag{6.5.76}$$

From this we immediately get

$$U^\dagger U = 1 \ . \tag{6.5.77}$$

And since $D_U = R_U = \mathcal{H}$ we also get

$$UU^\dagger = 1 \ . \tag{6.5.78}$$

Note, unlike the case for finite vector spaces, (6.5.77) does not imply (6.5.78) without the additional assumptions about domains on U. We now show that A is self-adjoint iff the operator

$$U = (A - i1)(A + i1)^{-1} \tag{6.5.79}$$

called the Cayley transform of A is unitary.

Proof
Suppose A is self-adjoint in $\mathcal{H}$ and $f \in D_A$. Then,

$$\begin{aligned} \| Af \pm if \|^2 &= (Af, Af) \pm i(Af, f) \mp i(f, Af) + (f, f) \\ &= \| Af \|^2 + \| f \|^2 \ . \end{aligned} \tag{6.5.80}$$

Therefore, $(A \pm i1)f = 0$ is only possible if $f = 0$. Thus, the operators $(A \pm i1)^{-1}$ and hence U exist. Furthermore, as we now show, the ranges R_{A+i1} and R_{A-i1} are dense in $\mathcal{H}$. For suppose g is orthogonal to all vectors in $R_{A \pm i1}$. Then for $f \in R_{A \pm i1}$ or equivalently, $f = (A \pm i1)h$ we have:

$$0 = (g, f) = (g, Ah \pm ih) = (g, Ah) \pm i(g, h) \ . \tag{6.5.81}$$

Thus,

$$(g, Ah) = \mp i(g, h) \ . \tag{6.5.82}$$

So,

$$g \in D_{A^\dagger} = D_A \tag{6.5.83}$$

and

$$A^\dagger g = Ag = \pm ig \ . \tag{6.5.84}$$

But as we have seen this is not possible unless $g = 0$. Thus, the ranges $R_{A \pm i1}$ are dense in $\mathcal{H}$. We now prove that in fact

$$R_{A \pm i1} = \mathcal{H} \ . \tag{6.5.85}$$

Let $g \in \mathcal{H}$, then since $R_{A \pm i1}$ is dense in $\mathcal{H}$ the limit $g_n = Af_n \pm if_n \to g$ exists. Also, using (6.5.80)

$$\begin{aligned} \| g_n - g_m \|^2 &= \| A(f_n - f_m) \pm i(f_n - f_m) \|^2 \\ &= \| A(f_n - f_m) \|^2 + \| (f_n - f_m) \|^2 \end{aligned} \tag{6.5.86}$$

and thus the f_n and Af_n converge to some vectors f and h respectively. Furthermore, because A is self-adjoint

$$f \in D_A \tag{6.5.87}$$

and

$$h = Af \ . \tag{6.5.88}$$

Hence, by definition of g as the limit of g_n, it follows that

$$g = Af \pm if \in R_{A\pm i1} \ . \tag{6.5.89}$$

Thus, the limit g of the approximating vectors g_n itself belongs to $R_{A\pm i1}$. However, by taking limits of sequences one may obtain any vector in $\mathcal{H}$. Thus, this limit may be any element in $\mathcal{H}$. Hence,

$$R_{A\pm i1} = \mathcal{H} \tag{6.5.90}$$

and we have that

$$D_U = R_U = \mathcal{H} \ . \tag{6.5.91}$$

Now choose any element f, then $f \in D_U$ and hence $f \in D_{(A\pm i1)^{-1}}$. Thus, we can write

$$f = (A + i1)g \tag{6.5.92}$$

and

$$\begin{aligned} Uf &= (A - i1)(A + i1)^{-1}(A + i1)g \\ &= (A - i1)g \ . \end{aligned} \tag{6.5.93}$$

Therefore,

$$\begin{aligned} \| Uf \|^2 &= \| (A - i1)g \|^2 = \| Ag \|^2 + \| g \|^2 \\ &= \| (A + i1)g \|^2 \\ &= \| f \|^2 \ . \end{aligned} \tag{6.5.94}$$

Thus, assuming $A = A^\dagger$ we conclude that $UU^\dagger = U^\dagger U = 1$. It is also always possible to recover A from U according to

$$A = i(1 - U)^{-1}(1 + U) = i(1 + U)(1 - U)^{-1} \ . \tag{6.5.95}$$

We now prove the converse, that if U is unitary, then A is self-adjoint. Let $g \in D_{A^\dagger}$ and define

$$\tilde{g} = A^\dagger g \ . \tag{6.5.96}$$

Then for any $f \in D_A$

$$(g, Af) = (\tilde{g}, f) \ . \tag{6.5.97}$$

But since

$$A = i(1 + U)(1 - U)^{-1} \ , \tag{6.5.98}$$

all $f \in D_A$ are of the form

$$f = (1-U)h \tag{6.5.99}$$

where

$$h \in D_U = \mathcal{H} \ . \tag{6.5.100}$$

Therefore, (6.5.97) reads:

$$(g, i(1+U)h) = (\tilde{g}, (1-U)h) \tag{6.5.101}$$

for any $h \in \mathcal{H}$. Now since U is unitary and therefore defined everywhere and conserves inner products, we can replace (h,g) by (Uh, Ug) and $(h,\tilde{g})$ by $(Uh, U\tilde{g})$ to get from (6.5.101)

$$(Ug, iUh) + (g, ih) - (U\tilde{g}, Uh) + (\tilde{g}, Uh) = 0 \tag{6.5.102}$$

or

$$(-iUg - ig - U\tilde{g} + \tilde{g}, Uh) = 0 \ . \tag{6.5.103}$$

Thus, $-iUg - ig - U\tilde{g} + \tilde{g}$ is orthogonal to all elements of $\mathcal{H}$ and hence vanishes

$$-iUg - ig - U\tilde{g} + \tilde{g} = 0 \ . \tag{6.5.104}$$

From this we get

$$g = -i\tilde{g} - U(g - i\tilde{g}) \ . \tag{6.5.105}$$

We now perform some algebra. Thus,

$$g = \frac{g - i\tilde{g}}{2} - \frac{g + i\tilde{g}}{2} - U(g - i\tilde{g}) \ . \tag{6.5.106}$$

Using (6.5.104) again this becomes

$$g = \frac{g - i\tilde{g}}{2} + \frac{U(g - i\tilde{g})}{2} - U(g - i\tilde{g}) \tag{6.5.107}$$

and hence

$$g = (1-U)\frac{g - i\tilde{g}}{2} \ . \tag{6.5.108}$$

Similarly we get

$$\tilde{g} = i(1+U)\frac{g - i\tilde{g}}{2} \ . \tag{6.5.109}$$

This proves two things: If $g \in D_{A^\dagger}$ then a) $g \in D_A$ according to (6.5.108), i.e. it is in $D_{(1-U)^{-1}}$.

$$\text{b) } Ag = \tilde{g} = A^\dagger g \tag{6.5.110}$$

since

$$Ag = i(1+U)(1-U)^{-1}(1-U)\frac{(g-i\tilde{g})}{2} \tag{6.5.111}$$

or

$$Ag = i(1+U)\frac{(g-i\tilde{g})}{2} = \tilde{g} \ . \tag{6.5.112}$$

This proves that $A^\dagger = A$ and hence that A is self-adjoint. Before proceeding let us examine the reasons for our interest in self-adjointness. The examination will be, of necessity, somewhat cursory.

6.6 Some Properties of Self-Adjoint Operators

To begin, consider the operator A which means multiplying by a number a. Self-adjointness implies that for $f, g \in \mathcal{H}$

$$(f, ag) = (af, g) \ . \tag{6.6.113}$$

But according to the definition of the inner product we have

$$(f, ag) = (a^* f, g) \ . \tag{6.6.114}$$

Thus,

$$a = a^* \ . \tag{6.6.115}$$

and hence a must be real. This is not a coincidence. In fact self-adjoint operators, in some sense which will become clear, correspond to real numbers. To make this precise we now discuss the eigenvalue problem for self-adjoint operators. Again, let A be a self-adjoint operator; then there are certain vectors belonging to the domain of A on which operations by A are particularly simple. The operation involves multiplication by a number. We have already encountered this in our solution of the Schrödinger equation

$$H\psi_E = E\psi_E \ . \tag{6.6.116}$$

Here, operating with H on the vector ψ_E involves multiplying ψ_E by E. The vector ψ_E is called an eigenfunction of H belonging to the eigenvalue E. More generally, f_j is an eigenfunction of the operator A belonging to the eigenvalue a_j if

$$Af_j = a_j f_j \ . \tag{6.6.117}$$

The important properties of self-adjoint operators are that:

a) All eigenvalues of a self-adjoint operator are real.

b) Eigenvectors belonging to different eigenvalues are orthogonal.

c) The eigenvectors form a complete set.

We now prove a) and b). The eigenvalues of a self-adjoint operator are real and the eigenvectors belonging to different eigenvalues are orthogonal.

Proof

Let f_i, f_j be two eigenfunctions of A belonging to the eigenvalues a_i and a_j respectively. Thus,

$$Af_i = a_i f_i \tag{6.6.118}$$

and

$$Af_j = a_j f_j \ . \tag{6.6.119}$$

Forming the inner product of (6.6.118) with f_i we get

$$(f_i, Af_i) = (Af_i, f_i) = (f_i, a_i f_i) = (a_i f_i, f_i) \ . \tag{6.6.120}$$

Therefore,

$$a_i = a_i^* \tag{6.6.121}$$

as required. Note the self-adjointness was necessary for otherwise we do not know that $f_i \in D_{A^\dagger}$. This proves part a).

To prove part b) we use (6.6.118) and (6.6.119) to get

$$(f_i, Af_j) = (Af_i, f_j) \ . \tag{6.6.122}$$

This means that

$$a_j(f_i, f_j) = a_i(f_i, f_j) \tag{6.6.123}$$

so that

$$(a_j - a_i)(f_i, f_j) = 0 \ . \tag{6.6.124}$$

Thus, if $a_i \neq a_j$, then

$$(f_i, f_j) = 0 \ . \tag{6.6.125}$$

Thus, we have established the results.

The proof of completeness of the eigenfunctions of a self-adjoint operator is beyond the scope of this book. Consequently we only show a sort of converse which makes the result appear plausible. The general theorem is known as the *Spectral Theorem* and is discussed in detail in reference [6.1].

Let A be a linear operator with a complete orthonormal set of eigenvectors $\{f_n\}$ and corresponding set of real eigenvalues a_n, then A is self-adjoint. Thus, we have that if

$$Af_n = a_n f_n \tag{6.6.126}$$

and

$$(f_n, f_m) = \delta_{nm} \tag{6.6.127}$$

where the $\{a_n\}$ are real and the $\{f_n\}$ are complete. Then $A = A^\dagger$.

Proof
We must show that $D_A = D_{A^\dagger}$ and that for $f, g \in D_A$

$$(Af, g) = (f, Ag) \, . \tag{6.6.128}$$

The proof is based on knowing A on a basis (the eigenfunctions). Suppose $f, g \in D_A$. Then because the $\{f_n\}$ are complete we have the following expansions

$$\begin{aligned} f &= \sum_n \alpha_n f_n \\ g &= \sum_n \beta_n f_n \end{aligned} \tag{6.6.129}$$

where

$$\begin{aligned} \alpha_n &= (f_n, f) \\ \beta_n &= (f_n, g) \, . \end{aligned} \tag{6.6.130}$$

Suppose $f \in D_A$. Then,

$$\begin{aligned} (f, Af) &= \sum_n (\alpha_m f_m, A\alpha_n f_n) \\ &= \sum_n a_n \alpha_m^* \alpha_n \\ &= \sum_n (a_m \alpha_m f_m, \alpha_n f_n) \\ &= \sum_n (A\alpha_m f_m, \alpha_n f_n) \\ &= (Af, f) = (A^\dagger f, f) \, . \end{aligned} \tag{6.6.131}$$

Therefore, $f \in D_{A^\dagger}$ and in a similar manner we get $(g, Af) = (Ag, f)$. Thus, A is self-adjoint.

6.7 Classification of Symmetric Operators

We now complete the classification of symmetric operators. For the purposes of physics there is no need to distinguish between self-adjoint and essentially self- adjoint operators since the latter always have a unique and obvious extension to self-adjoint operators. We are mainly concerned in determining which symmetric operators have several self-adjoint extensions. Our main tool in this investigation is the Cayley transform that we discussed previously.

Suppose A is an arbitrary symmetric operator (not necessarily self-adjoint). In that case the corresponding Cayley transform U need not be unitary and the domain D_U and the range $R_U = UD_U$ need not coincide with the whole Hilbert space $\mathcal{H}$. If we consider the sets of vectors $D_U^\perp$ and $R_U^\perp$ orthogonal to D_U and R_U, the "size" of these sets gives us an indication of the extent to which U is not unitary and A is not self-adjoint. It is straightforward to check that the sets $D_U^\perp$ and $R_U^\perp$ are in fact subspaces. We call these subspaces the *deficiency subspaces* of A and their dimensions the *deficiency indices*. Thus, the deficiency indices of A are

$$(m, n) = (\dim D_U^\perp , \dim R_U^\perp) . \tag{6.7.132}$$

Now all elements in D_U are of the form

$$f = (A + i1)g . \tag{6.7.133}$$

Thus, if $h \in D_U^\perp$, then

$$(h, f) = (h, (A + i1)g) = 0 \tag{6.7.134}$$

or

$$(h, Ag) = -i(h, g) = (ih, g) . \tag{6.7.135}$$

So $h \in D_{A^\dagger}$ and

$$A^\dagger h = ih . \tag{6.7.136}$$

But, if this is true, then,

$$(h, (A + i1)g) = ((A^\dagger - i1)h, g) = 0 \tag{6.7.137}$$

and $h \in D_U^\perp$. So, we have shown that $h \in D_U^\perp$ iff

$$A^\dagger h = ih \tag{6.7.138}$$

for any $h \in D_{A^\dagger}$.

Thus, $\dim D_U^\perp$ is given by the number of linearly independent solutions of

$$A^\dagger h = ih \tag{6.7.139}$$

belonging to $D_{A^\dagger}$.

Similarly, all vectors in R_U are of the form $(A - i1)g$ and hence by a set of steps like above, we find that $\dim R_U^\perp$ is given by the number of linearly independent solutions of

$$A^\dagger h = -ih \tag{6.7.140}$$

belonging to $D_{A^\dagger}$.

Now from our previous results we know that A is self-adjoint iff U is unitary and hence iff $R_U = D_U = \mathcal{H}$. Thus, A is self-adjoint iff the deficiency indices are $(0, 0)$. To see what this means we return to our previous example of the

momentum operator defined on the interval (a, b) and begin with the symmetric, but not self-adjoint operator with domain

$$D_p = \{f \in C^1 \mid f(a) = f(b) = 0\} \tag{6.7.141}$$

where C^1 means "functions whose first derivative is continuous". If we now consider the expression

$$\begin{aligned}(h, pf) &= \int_a^b h^* \frac{\hbar}{i}\frac{df}{dx}\,dx \\ &= \frac{\hbar}{i} h^* f|_a^b + \int_a^b \left(\frac{\hbar}{i}\frac{dh}{dx}\right)^* f\,dx\end{aligned} \tag{6.7.142}$$

it follows from $f(a) = f(b) = 0$ that $h \in D_{p^\dagger}$ for any $h \in \mathcal{H}$ such that both $h, h' \in \mathcal{L}_2$. Furthermore, the action of $p^\dagger$ is the same as that of p, namely $(\hbar/i)\ (d/dx)$.

The only solutions of (6.7.136) and (6.7.140) respectively are:

$$h_+ = Ae^{-x/\hbar}\ ,\ h_- = Ae^{x/\hbar}\ . \tag{6.7.143}$$

Thus, the deficiency indices are (1,1).

In general if the deficiency indices of a symmetric operator A are (m, n) with $(m \neq 0, n \neq 0)$ it is possible to extend (increase the domain) the operator A as follows. Let two solutions of (6.7.136) and (6.7.140) be h_+ and h_- respectively. Then for $g \in D_A$ define the operator A' by:

$$A'[g + \theta(h_+ + h_-)] = Ag + i\theta(h_+ - h_-)\ . \tag{6.7.144}$$

Clearly A' is an extension of A since now $h_+ + h_-)$ belongs to $D_{A'}$. Furthermore, the deficiency indices for A' are $(m-1, n-1)$. To see this one simply needs to verify that the corresponding Cayley transform of A, namely U is extended to U' where

$$U'f = Uf \text{ if } f \in D_U \tag{6.7.145}$$

and

$$U'h_+ = h_-\ . \tag{6.7.146}$$

One can proceed in this manner until one gets deficiency indices $(r, 0)$ or $(0, r)$. In this case no further extension is possible. If one then finds that $r \neq 0$, one has to conclude that such an operator has no self-adjoint extensions.

Although the procedure above yields the self-adjoint extensions of symmetric operators with deficiency indices (n, n) it is not the most useful approach. For physicists boundary conditions are usually of important physical significance with direct physical interpretations. This is brought out in the approach we now take. For example, if the deficiency indices are $(1, 1)$ we get a one-parameter family of self-adjoint extensions, and for deficiency indices (n, n) we get an n^2-parameter family of self-adjoint extensions. Mathematically this is as far as one can go. To pick the "correct" extension in these cases depends on the

physical situation and cannot be decided by mathematics. It requires physics. To illustrate these points we now discuss some specific examples.

Again consider the momentum operator $p = (\hbar/i)\,(d/dx)$ on $\mathcal{L}_2$. Let p be defined on the interval $(-\infty, \infty)$. In this case the equations for the deficiency indices read as before

$$\frac{\hbar}{i}\frac{df}{dx} = \pm if \tag{6.7.147}$$

or

$$\frac{df}{dx} = \mp\frac{1}{\hbar}f\,. \tag{6.7.148}$$

The solutions are

$$f = Ae^{\mp x/\hbar}\,. \tag{6.7.149}$$

But neither of these solutions is square integrable on $(-\infty, \infty)$ and hence the deficiency indices are $(0,0)$ and defined over the interval $(-\infty < x < \infty)$ the operator p is self-adjoint.

Now consider the case where the operator $(\hbar/i)\,(d/dx)$ is defined on the interval $(0, \infty)$. As before the solutions are

$$f = Ae^{\mp x/\hbar}\,. \tag{6.7.150}$$

This time

$$f = Ae^{-x/\hbar} \tag{6.7.151}$$

is square integrable, but

$$f = Ae^{x/\hbar} \tag{6.7.152}$$

is not. Thus, the deficiency indices are $(1,0)$ and defined over $(0 \le x < \infty)$ the operator $(\hbar/i)\,(d/dx)$ has no self-adjoint extensions. The reason for this is easy to see. Consider

$$\left(f, \frac{\hbar}{i}\frac{dg}{dx}\right) = \int_0^\infty (f^*\frac{\hbar}{i}\frac{dg}{dx})\,dx\,. \tag{6.7.153}$$

After integrating by parts, this becomes

$$\begin{aligned}\left(f, \frac{\hbar}{i}\frac{dg}{dx}\right) &= \frac{\hbar}{i}\,f^*g|_0^\infty + \int_0^\infty \left(\frac{\hbar}{i}\frac{df}{dx}\right)^* g\,dx \\ &= \frac{\hbar}{i}\,f^*g|_0^\infty + \left(\frac{\hbar}{i}\frac{df}{dx}, g\right)\,. \end{aligned}\tag{6.7.154}$$

So, for $(\hbar/i)\,(d/dx)$ to be self-adjoint requires that

$$\frac{\hbar}{i}\,f^*g|_0^\infty = 0\,. \tag{6.7.155}$$

Now we know that both f and g vanish for $x \to \infty$. Therefore we require

$$f(0) = 0 \text{ or } g(0) = 0 \,. \tag{6.7.156}$$

In fact to make $p = (\hbar/i)(d/dx)$ self-adjoint requires $D_p = D_{p\dagger}$ and hence both $f(0) = g(0) = 0$. On the other hand, the "eigenfunctions" of this p are Ae^{ikx} and only vanish at $x = 0$ if $A = 0$. Thus, this p would have no eigenfunctions if it were to be self-adjoint. There is also a physical reason for the lack of a self-adjoint extension in this case. We discuss this, in more detail, a little further on.

Finally, consider the operator $p = (\hbar/i)(d/dx)$ defined on the interval $a \leq x \leq b$. In this case, both solutions $Ae^{\pm x/\hbar}$ are square integrable and the deficiency indices are $(1,1)$. This means that the self-adjoint extensions should depend on precisely one parameter as we already indicated. We now examine what this parameter is. In this case we have

$$\begin{aligned}(f, pg) &= \int_a^b f^* \frac{\hbar}{i}\frac{dg}{dx}\,dx \\ &= \frac{\hbar}{i} f^* g\Big|_a^b + (pf, g) \,.\end{aligned} \tag{6.7.157}$$

Therefore for self-adjointness we require

$$f^*(b)g(b) = f^*(a)g(a) \tag{6.7.158}$$

or

$$\left(\frac{f(b)}{f(a)}\right)^* = \frac{g(a)}{g(b)} \,. \tag{6.7.159}$$

This implies that

$$\frac{f(b)}{f(a)} = \frac{g(b)}{g(a)} = e^{i\theta} \,. \tag{6.7.160}$$

So θ is the parameter determining the different self-adjoint extensions. Actually, this is a specification of boundary conditions. Thus, if we choose $\theta = 0$ we have periodic boundary conditions. These are the most common. Note that once we have specified that the domain of p is the set of all square-integrable functions for $a \leq x \leq b$ (abbreviated $\mathcal{L}_2(a,b)$) such that

$$f(b) = e^{i\theta} f(a) \,, \tag{6.7.161}$$

then the domain of $p^\dagger$ is the same as that of p and the deficiency indices become $(0,0)$ showing that p is self-adjoint. To give some idea of the physics behind these three situations we state a theorem without proof and then use it to explain the physics behind all this.

Stone's Theorem

If A is a self-adjoint operator, then

$$U(s) = e^{isA} \tag{6.7.162}$$

is a unitary operator for every real number s. Furthermore,

$$U(s)U(t) = U(s+t) \tag{6.7.163}$$

and

$$U(-s) = U(s)^\dagger = U(s)^{-1} \ . \tag{6.7.164}$$

Conversely given a set of continuous unitary operators satisfying (6.7.163) and (6.7.164) then there is a self-adjoint operator A such that (6.7.162) holds and A is given by

$$iA = \lim_{\epsilon \to 0} \frac{U(\epsilon) - 1}{\epsilon} \ . \tag{6.7.165}$$

Operators satisfying (6.7.163) and (6.7.164) form an algebraic structure called a group. We now construct such a set of unitary operators and use them to interpret what we did.

Consider the operator

$$U(s)f(x) = f(x+s) \ . \tag{6.7.166}$$

Then,

$$U(0)f(x) = f(x) \tag{6.7.167}$$

and

$$U(t)U(s)f(x) = U(t)f(x+s) = f(x+s+t) = U(t+s)f(x) \ . \tag{6.7.168}$$

Thus, the operators defined by (6.7.166) satisfy (6.7.163) and also (6.7.164) as we see by setting $t = -s$ for a left inverse and $s = -t$ for a right inverse. On the other hand, if $f(z)$ is analytic for $a \leq \Re(z) \leq b$ it has the Taylor expansion

$$\begin{aligned} f(x+s) &= \sum_n^\infty \frac{s^n}{n!} \frac{d^n}{dx^n} f(x) \\ &= \sum_n^\infty \left(\frac{isp}{\hbar}\right)^n \frac{1}{n!} f(x) \\ &= e^{isp/\hbar} f(x) \ . \end{aligned} \tag{6.7.169}$$

Thus,

$$U(s) = e^{isp/\hbar} \tag{6.7.170}$$

and we have succeeded in expressing $U(s)$ in the form (6.7.162). It makes sense to call $U(s)$ the translation operator since it "translates" functions by an amount s. Furthermore we call p the *generator* of translations since for infinitesimal translations.

$$U(s) \to 1 + i\frac{s}{\hbar} p \quad \text{as} \quad s/\hbar \to 0 \ . \tag{6.7.171}$$

In terms of these considerations we can understand why p has a one-parameter family of self-adjoint extensions on $\mathcal{L}_2(a,b)$. Actually, our considerations depend on the fact that f is analytic for $a \leq \Re(z) \leq b$. This, however, is not a restriction since the functions square-integrable and analytic over $a \leq \Re(z) \leq b$ form a dense set in $\mathcal{L}_2(a,b)$.

Suppose $f(x)$ is an infinitely differentiable function which is non-zero only on an interval completely contained in $a \leq x \leq b$. An example of such a function is

$$f(x) = \begin{cases} 0 & \text{if } x \leq \alpha < a \ , \ x \geq \beta > b \\ \exp\left(-\frac{1}{x-\alpha} - \frac{1}{\beta - x}\right) & \text{if } \alpha \leq x \leq \beta \end{cases} \ . \tag{6.7.172}$$

Note that this function is definitely not analytic over $a \leq x \leq b$ since it has essential singularities at both $x = \alpha$ and $x = \beta$. Also, an analytic function cannot vanish on an open interval unless it is identically zero. Now consider the normalization (i.e. unit probability) associated with f, namely, $\int_a^b |f(x)|^2 \, dx$. We want the translation operator $U(s)$ to preserve this normalization, that is, to be unitary. Thus, we need

$$\int_a^b |U(s)f(x)|^2 \, dx = \int_a^b |f(x)|^2 \, dx \ . \tag{6.7.173}$$

But,

$$U(s)f(x) = f(x+s) \tag{6.7.174}$$

and if $s > b - \beta$ part of the wavefunction "disappears" past the right end point. To conserve the integral above requires that what disappears at the right must reappear from the left. Of course the phase of the function can be shifted in reappearing from the left. Furthermore, all functions must experience the same phase shift. Thus, if f_1 and f_2 are two such functions and if their phase shifts are different, say θ_1 and θ_2, then translation of the function $f = f_1 + f_2$ will eventually produce the function $f' = e^{i\theta_1} f_1 + e^{i\theta_2} f_2$. But,

$$\int_a^b |U(s)f(x)|^2 \, dx \neq \int_a^b |f(x)|^2 \, dx \tag{6.7.175}$$

unless $\theta_1 = \theta_2$. So the superposition principle limits the number of phase shift parameters to one.

Why then does p not have any self-adjoint extensions on $\mathcal{L}_2(0,\infty)$? The answer is as follows. Translating x to the right will never take the function past the right endpoint. On the other hand, by translating to the left we can always bring the function past the left endpoint (the origin). In this case there is not anywhere from where the function can reappear to conserve probability and hence p can not be made self-adjoint. This also explains why p is already self-adjoint on $\mathcal{L}_2(-\infty,\infty)$.

Another extremely simple problem is the case of a particle in a strongly repulsive potential such as a quartic or cubic potential

$$V(x) = -gx^n \quad g > 0 \quad n = 3,4 \ . \tag{6.7.176}$$

In this case the Hamiltonian is

$$H = \frac{p^2}{2m} + V(x) = \frac{p^2}{2m} - gx^n \,. \tag{6.7.177}$$

By redefining the variables we can bring this to the form

$$H = -\frac{d^2}{dx^2} - ax^n \quad a > 0 \,. \tag{6.7.178}$$

We want to examine this Hamiltonian on $\mathcal{L}_2(-\infty, \infty)$. To get a feel for the physical situation consider the problem classically. In the previous example we saw that the existence of different self-adjoint extensions depended on the fact that the particle can reach a boundary (a or b) and have to be transmitted or reflected.

In this case the boundaries are at $\pm\infty$. So we must see if the particle can in fact reach these boundaries. Now suppose the particle starts at $x = 0$ with energy $E > 0$. Then classically its velocity v is given by

$$\frac{1}{2}mv^2 - gx^n = E \tag{6.7.179}$$

or

$$v = \sqrt{\frac{2E}{m} + \frac{2}{m}gx^n} \,. \tag{6.7.180}$$

Therefore the time to reach ∞ is

$$t = \int_0^\infty \frac{dx}{v} = \int_0^\infty \frac{dx}{\sqrt{\frac{2E}{m} + \frac{2}{m}gx^n}} \,. \tag{6.7.181}$$

And for $n = 3$ or 4, $t < \infty$. So, the particle reaches $+\infty$ in a finite time. To conserve probability it must be reflected and return to the origin in a finite time. Thus, the time-averaged particle position is near the origin. We therefore expect to find that all eigenfunctions of this Hamiltonian are square integrable and that the spectra of the self-adjoint extensions of H are discrete. This is, in fact, the case and this Hamiltonian is analogous to a free particle Hamiltonian on a finite finite interval (a, b), (see problem 6.7). For the repulsive quartic potential the points $\pm\infty$ behave like the end points (a, b). So it is not surprising that the deficiency indices are $(2, 2)$ for both cases. For the repulsive cubic potential the situation is different. The particle can again reach $x = +\infty$ in a finite time but it can never reach $x = -\infty$. Thus we need only specify boundary conditions at $x = +\infty$. In this case the deficiency indices turn out to be $(1, 1)$. The main point of this discussion is that whenever an operator, which is a candidate for representing an observable, is not self-adjoint but has self-adjoint extensions, then there are good physical reasons for this.

6.8 Spontaneously Broken Symmetry

The concept of spontaneously broken symmetries plays a very important role in some field theories of elementary particles. Since it fits naturally into the topics we have just discussed, we shall start by defining the concept and then proceed to illustrate it with an example.

Suppose we have some observable, whose representative operator Q commutes with the Hamiltonian H. Then either Q corresponds to a discrete symmetry operation Q_D such as parity, or else we can use Q to define a one-parameter family of unitary operators,

$$U(\alpha) = e^{i\alpha Q} \cdot \tag{6.8.182}$$

for which Q is the generator. It then follows from

$$[Q, H] = 0 \tag{6.8.183}$$

that

$$e^{i\alpha Q} H e^{-i\alpha Q} = H \tag{6.8.184}$$

or else for a discrete symmetry

$$Q_D H Q_D^{-1} = H \, . \tag{6.8.185}$$

We also require that the ground state of the Hamiltonian ϕ_0 should be invariant under either $U(\alpha)$ or Q_D. This means

$$U(\alpha)\phi_0 = \phi_0 \tag{6.8.186}$$

or

$$Q_D \phi_0 = \phi_0 \, . \tag{6.8.187}$$

Definition
A symmetry corresponding to an observable Q (Q_D) is spontaneously broken if all the above statements except (6.8.186) (respectively (6.8.187)) hold. For this to occur requires that the ground state be degenerate. This, in itself, is an unusual phenomenon. To illustrate this phenomenon we consider the Hamiltonian

$$H = \frac{p^2}{2m} \tag{6.8.188}$$

defined on the interval $-a \leq x \leq a$. We furthermore pick for $p = (\hbar/i)\,(d/dx)$ the self-adjoint extension corresponding to the domain

$$D_p = \left\{ f(x) \in C^1 \mid f(a) = -f(-a) \right\} \, . \tag{6.8.189}$$

Thus, instead of periodic we pick "anti-periodic" boundary conditions.

The complete set of normalized eigenfunctions of this momentum operator are given by:

$$f_n(x) = \frac{1}{\sqrt{2a}} e^{i\pi(n+1/2)x/a} \quad n = 0, \pm 1, \pm 2, \ldots \tag{6.8.190}$$

with corresponding eigenvalues $(n + 1/2)\pi\hbar/a$. These wavefunctions have the following symmetry properties

$$f_n(x) = f_{-(n+1)}(-x) \tag{6.8.191}$$

and

$$f_n^*(x) = f_{-(n+1)}(x) \, . \tag{6.8.192}$$

Thus, the parity operator P and the time-reversal operator T have the following action on them

$$(Pf_n)(x) = f_{-(n+1)}(x) \tag{6.8.193}$$

$$T(f_n)(x) = f_{-(n+1)}(x) \, . \tag{6.8.194}$$

The set of functions $\{f_n\}$ are also eigenfunctions of the Hamiltonian. In fact,

$$Hf_n = \frac{\pi^2\hbar^2}{2ma^2}(n + 1/2)^2 f_n \tag{6.8.195}$$

$$Hf_{-(n+1)} = \frac{\pi^2\hbar^2}{2ma^2}(n + 1/2)^2 f_{-(n+1)} \, . \tag{6.8.196}$$

Thus, all eigenvalues including the ground state eigenvalue

$$E_0 = E_{-1} = (\pi^2\hbar^2)/(2ma^2)$$

are *doubly degenerate.* We further see that although the Hamiltonian H, the parity operator P, and the time-reversal operator T commute, the two ground states f_0 and f_{-1} are not eigenstates of either the parity operator or the time-reversal operator

$$(Pf_0)(x) = T(f_0)(x) = f_{-1}(x) \tag{6.8.197}$$

and

$$(Pf_{-1})(x) = T(f_{-1})(x) = f_0(x) \, . \tag{6.8.198}$$

Thus, parity and time-reversal are spontaneously broken symmetries.

It is possible to restore these symmetries by defining states

$$g_n^+(x) = \frac{1}{\sqrt{2}}\left[f_n(x) + f_{-(n+1)}(x)\right] = \frac{1}{\sqrt{a}} \cos(n + 1/2)\frac{\pi x}{a} \tag{6.8.199}$$

as well as

$$g_n^-(x) = \frac{1}{\sqrt{2}}\left[-f_n(x) + f_{-(n+1)}(x)\right] = \frac{1}{\sqrt{a}} \sin(n + 1/2)\frac{\pi x}{a} \, . \tag{6.8.200}$$

These are now simultaneous eigenstates of H, P, and T.

$$(Pg_n)^{\pm}(x) = \pm\, g_n^{\pm}(x) \tag{6.8.201}$$

$$(Tg_n)^{\pm}(x) = \pm\, g_n^{\pm}(x) \tag{6.8.202}$$

$$Hg_n^{\pm}(x) = \frac{\pi^2\hbar^2}{2ma^2}\, g_n^{\pm}(x)\ . \tag{6.8.203}$$

So, parity and time reversal are no longer broken symmetries. In this case, however, we have an even more surprising symmetry breaking, for although the momentum operator p and the Hamiltonian $p^2/2m$ commute, these eigenstates of the Hamiltonian are not eigenstates of the momentum operator. In fact,

$$pg_0^{\pm}(x) = \mp\frac{\pi\hbar}{a}(n+1/2)\, g_n^{\mp}(x) \tag{6.8.204}$$

so that in particular

$$pg_0^{+}(x) = -\frac{\pi\hbar}{2a}\, g_0^{-}(x) \tag{6.8.205}$$

and

$$pg_0^{-}(x) = +\frac{\pi\hbar}{2a}\, g_0^{+}(x)\ . \tag{6.8.206}$$

In this case we therefore have translational symmetry spontaneously broken since the translation operator

$$U(s) = e^{isp/\hbar} \tag{6.8.207}$$

does not leave the ground states $g_0^{\pm}$ invariant. In fact, by expanding

$$U(s) = \sum_n \frac{1}{n!}\left(\frac{isp}{\hbar}\right)^n \tag{6.8.208}$$

and repeatedly applying (6.8.205) and (6.8.206) we get:

$$U(s)g_0^{\pm}(x) = \cos\left(\frac{\pi s}{2a}\right)\, g_0^{\pm}(x) \mp i\sin\left(\frac{\pi s}{2a}\right)\, g_0^{\mp}(x)\ . \tag{6.8.209}$$

This demonstrates conclusively that the translational symmetry is broken.

We now relate the mathematical model we have displayed, to a definite physical system. If one considers a one-dimensional crystal consisting of only one type of atom, then the boundary condition in going from nearest neighbour to nearest neighbour is periodic. The situation repeats itself. Similarly for a one-dimensional crystal with alternating atoms $(ABAB)\ldots$, as in an antiferromagnet, the boundary condition from an atom to its next nearest neighbour is periodic, and hence from nearest neighbour to nearest neighbour anti-periodic.

We can now visualize the physical situation corresponding to our model and get a clearer understanding of the cause of the broken symmetry. If we consider such an antiferromagnetic crystal and consider the interval between nearest

neighbours as fundamental, we must impose antiperiodic boundary conditions. Furthermore, since the end points correspond physically to different situations (atoms) it makes a difference whether a particle travels freely from left to right or right to left. The situations are not mirror images of each other and hence not eigenstates of the parity operator. Since time-reversal reverses the direction of travel, these states are also not eigenstates of the time-reversal operator.

One can take a superposition of states of particles travelling to the left and right, as we did, to get standing waves which are then automatically time-reversal as well as parity invariant. In this case, however, conservation of probability brings about a loss of translation invariance. It is clear now that this "unusual" self-adjoint extension of the momentum operator has just as physical an interpretation as the usual one with periodic boundary conditions.

It is perhaps also worth while to notice that the commutation relation

$$[x, p] = i\hbar \tag{6.8.210}$$

is not valid in this representation since for $f \in D_p$, $xf \notin D_p$ in general. In fact, in this case, $xf \in D_p$ only if $f(a) = f(-a) = 0$. Nevertheless, it is true that

$$[x^{2n}, p] = i2n\hbar\, x^{2n-1} \quad n = 0, 1, 2, \ldots \ . \tag{6.8.211}$$

This concludes our mathematical treatment of self-adjointness. We now turn to a systematic analysis of the physical interpretation of quantum mechanics.

6.9 Problems

6.1 Consider the set of functions $\{f_k(x) = e^{ikx} f(x)\ ,\ f(x) \in \mathcal{L}_2\}$. Show that

$$\lim_{k\to\infty} (g, f_k) = 0 \quad g \in \mathcal{L}_2$$

whereas

$$\| f_k \|^2 = \| f \|^2 \neq 0 \ .$$

The above type of convergence of $f_k \to 0$ is called weak convergence as opposed to the notion of strong convergence defined in the text. Hint: Use the Riemann-Lebesgue Theorem [1].

6.2 Show that every Cauchy sequence in a finite dimensional vector space converges strongly (see problem 6.1).

6.3 Consider the operator $A = px^{2n+1} + x^{2n+1}p$ where $p = (\hbar/i)(d/dx)$ and $n = 1, 2, 3, \ldots$. Find the eigenvalues and eigenfunctions of A. What are the deficiency indices of A ? The Hilbert space in this case is $\mathcal{L}_2(-\infty, \infty)$. For $n = 1$ this example is due to von Neumann.

6.4 A projection operator is a self-adjoint, non-negative operator P satisfying $P^2 = P$. Let f_n be a normalized eigenfunction of a self-adjoint operator A with only discrete eigenvalues λ_n .
a) Show that the operator $P_n\psi = f_n(f_n, \psi)$ is a projection operator.
b) Show that A can be written

$$A\psi = \sum_n \int \lambda_n f_n(x) f_n^*(y)\psi(y)\, dy = \sum_n \lambda_n P_n \psi \ .$$

This is called the spectral resolution of the operator A.
Hint: Assume completeness of the eigenfunctions.

6.5 Find the spectral resolution (see problem 6.4) of the operator

$$A = \begin{pmatrix} a_3 & a_1 + ia_2 \\ a_1 - ia_2 & -a_3 \end{pmatrix} \quad a_i \text{ real.}$$

6.6 For any operator A the corresponding operator $R(z) = (A - z1)^{-1}$, where 1 stands for the unit operator, is called the resolvent operator. Show that for any square matrix A , $R(z)$ is analytic in z with poles at the eigenvalues of A.

6.7 Find the deficiency indices and hence all self-adjoint extensions of the Hamiltonian

$$H = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}$$

defined on the interval (a, b).
Hint: It may be useful to express the boundary conditions on a function $f \in D_H$ in terms of 2-component quantities

$$F(a) = \begin{pmatrix} f(a) \\ f'(a) \end{pmatrix} \quad \text{and} \quad F(b) = \begin{pmatrix} f(b) \\ f'(b) \end{pmatrix}$$

and assume that $F(b) = UF(a)$ where U is a non-singular 2×2 matrix.

6.8 Given an orthonormal basis set $\{u_n : n = 0, 1, 2, \ldots\}$ and an operator a which has the following action on this basis:

$$a u_n = \sqrt{n}\, u_{n-1} \quad n \geq 0 \ .$$

Find the adjoint operator $a^\dagger$ by explicitly giving its action on this basis set. Also find the commutator $[a, a^\dagger]$.

Bibliography

[6.1] G.F. Simmons, *Introduction to Topology and Modern Analysis* - McGraw-Hill Book Co. (1963) Chapter 10.

[6.2] A.S. Wightman, Chapter 8 in *Cargese Lectures in Theoretical Physics*, Edited by M. Levy - Gordon and Breach Inc. (1967).

[6.3] A.H. Zemanian, *Distribution Theory and Transform Analysis* - McGraw-Hill Book Co. (1965) page 174.

[6.4] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* - Princeton University Press (1955).

[6.5] T. Kato, *Perturbation Theory for Linear Operators* - Springer-Verlag, New York Inc. (1966). This book gives a very concise and complete summary of useful Hilbert space techniques.

[6.6] Two of the standard treatises on functional analysis are:
F. Riesz and B.S.Z. Nagy, *Functional Analysis* - Frederick Unger Publishing Co., N.Y. (1955) and
K. Yosida - *Functional Analysis* - Springer-Verlag, New York Inc. (1966).