CHAPTER 5

VECTOR SPACES

A vector space is a set (of elements) with certain special properties

Vector spaces are one of the most fundamental mathematical structures in physical theory.

Vector spaces are far more general than vectors in ordinary space

Basically, this subject deals with *quantities (any arbitrary function)* that can be represented *by expansions in a series of (known basis) functions*, and includes the *methods* by which such *expansions can be generated and used* for various purposes.

A key aspect of the subject is the notion that a more or less arbitrary **function** can be represented by such an expansion. For example we have a two dimensional space in which the two coordinates which are real numbers that we will call a_1 and a_2 are associated with the two (basis) functions $\varphi_1(s)$ and $\varphi_2(s)$. In this space the coordinate points (a_1 , a_2) correspond to the function f(s) as

$$f(s) = a_1\, \varphi_1(s) + a_2\, \varphi_2(s). \qquad (5.1)$$

The analogy with a physical two dimensional physical space with vectors

$\vec{A} = A_i\, \hat{\imath} + A_j\, \hat{\jmath}$ is that while $f(s) \leftrightarrow \vec{A}$,

$\varphi_1(s)$ correspond to $\hat{\imath}$ and $a_1 \leftrightarrow A_i$,

$\varphi_2(s)$ correspond to $\hat{\jmath}$ and $a_2 \leftrightarrow A_j$

Equation (5.1) defines a set of functions (a **function space**) that can be built from the **basis** φ_1, φ_2; we call this space a **linear vector space** because its members are linear combinations of the basis functions and the addition of its members corresponds to component (coefficient) addition. If $f(s)$ is given by Eq. (5.1) and $g(s)$ is given by another linear combination of **the same** basis functions,

$$g(s) = b_1\varphi_1(s) + b_2\varphi_2(s),$$

with b_1 and b_2 the coefficients defining $g(s)$, then

$$h(s) = f(s) + g(s) = (a_1 + b_1)\varphi_1(s) + (a_2 + b_2)\varphi_2(s) \qquad (5.2)$$

defines $h(s)$, the member of our space (i.e., the function), which is the sum of the members $f(s)$ and $g(s)$. In order for our vector space to be useful, we consider only spaces in which the sum of any two members of the space is also a member.

In addition, the notion of linearity includes the requirement that if $f(s)$ is a member of our vector space, then $u(s) = k\, f(s)$, where k is a real or complex number, is also a member, and we can write

$$u(s) = k\, f(s) = ka_1\varphi_1(s) + ka_2\varphi_2(s). \qquad (5.3)$$

Vector spaces for which addition of two members or multiplication of a member by scalar always produces a result that is also a member are termed **closed** under these operations.

- Vector spaces for which addition A+B of two members A, B always produces a result C that is also a member
- or multiplication of a member A by scalar k always produces a result kA = B that is also a member
- Vector spaces are termed **closed** under those operations -> addition (+) and multiplication ($\cdot$)
- multiplication of f by a ordinary number k (which, by analogy with ordinary vectors, we call a **scalar**), results in the multiplication of the coefficients by k.
- These are exactly the operations we would carry out to form the sum of two ordinary vectors, **A**+ **B**, or the multiplication of a vector by a scalar, as in k**A**. However, here we have the coefficients a_i and b_i , which combine under vector addition and multiplication by a scalar in exactly the same way that we would combine the ordinary vector components A_i and B_i

Basis (functions) Vectors

- The main universal restriction on the form of a basis is that the
- basis members be linearly independent, so that
- any function (member) of our vector space will be described by a unique linear combination of the basis functions.

The number of basis functions (i.e., the dimension of our basis) may be a small number such as 2, 3, a large but finite integer or even denumerably infinite (as would arise in an untruncated power series). The main universal restriction on the form of a basis is that the basis members be linearly independent, so that any function (member) of our vector space will be described by a unique linear combination of the basis functions. We illustrate the possibilities with some simple example.

Example: SOME VECTOR SPACES

1. We consider first a *vector space* of dimension 3, which is **spanned by** (meaning that it has a basis set S that consists of) the following three functions

$$S = \{ P_o(s) = 1 ,\ P_1(s) = s,\ P_2(s) = (\tfrac{3}{2}s^2 - \tfrac{1}{2}) \} \qquad (5.4)$$

Express the following function in terms of the the above basis (functions) $P_o(s)$, $P_1(s)$ and $P_2(s)$

(i) s+3 (ii) s^2 (iii) 4-3s

Solution: (i) s + 3 can be written as s + 3. **1** ➔ s + 3 = $P_1(s)$+ 3 $P_o(s)$

(i) To solve for s^2 let us consider $(\frac{3}{2}s^2 - \frac{1}{2}) = P_2(s)$

From which we get $\frac{3}{2}s^2 = \frac{1}{2} + P_2(s)$

$$s^2 = \frac{1}{3}\,.\,\mathbf{1} + \frac{2}{3}P_2(s)$$

➔ $$s^2 = \frac{1}{3} P_o(s) + \frac{2}{3}P_2(s)$$

(iii) 4-3s can be written as 4 .**1** - 3 .s ➔ 4-3s = 4 $P_o(s)$ - 3 $P_1(s)$

In fact, because we can write 1, s, and s^2 in terms of our basis, we can see that **any** quadratic function in s will be a member of our vector space, and that our space includes only functions of s that can be written in the form $a_o + a_1 s + a_2 s^2$.

To illustrate our vector-space operations, for example:

we can write the polynomial $3 + 2s + 5s^2$ in terms of the basis functions $P_i(s)$ contained in set S as

$$3 P_o(s) + 2 P_1(s) + 5 P_2(s).$$

Important Note : We are free to define our basis any way we want, so long as ***its members are linearly independent***. We could have chosen another basis for this same vector space as

$$S_1 = \{ P_o(s) = 1 , P_1(s) = s, P_2(s) = s^2 \} \qquad (5.5)$$

Assignment : Express the functions (i) $s+3$ (ii) s^2 (iii) $4-3s$

in terms of the basis functions contained in set, $S_1 = \{ P_o(s) = 1 , P_1(s) = s, P_2(s) = s^2 \}$.

SCALAR PRODUCT

To make the vector space concept useful and parallel to that of vector algebra in ordinary space, we need to introduce the concept of a scalar product in our function space. We shall write the scalar product of two members of our vector space, f and g, as $\langle f|g \rangle$. This is the notation that is almost universally used in physics; various other notations can be found in the mathematics literature; examples include $[f; g]$ and $(f; g)$.

There exists an extremely wide range of possibilities for defining scalar products that meet these criteria. The situation that arises most often in physics is that the members of our vector space are ordinary functions of the variable s and the scalar product of the two members $f(s)$ and $g(s)$is computed as an integral of the type

$$\langle f \,|\, g \rangle = \int_a^b f^*(s) g(s) w(s) ds$$

with the choice of a, b, and $w(s)$ dependent on the particular definition we wish to adopt for our scalar product. In the special case $g = f$, the scalar product

$$\langle f|f \rangle = \int_a^b f^*(s) f(s) w(s) ds$$

is to be interpreted as the square of a “length,” and this scalar product must therefore be positive for any f zero only when f=0.

HILBERT SPACES

These are the vector spaces of primary importance in physics.

A Hilbert space H has the following properties:

- Elements (members) f, g, or h of $\mathcal{H}$ are subject to two operations, **addition**, and **multiplication by a scalar** (here k, k_1, or k_2). These operations produce quantities that are also members of the space.
- Addition is commutative and associative:

$$f(s) + g(s) = g(s) + f(s), \quad [f(s) + g(s)] + h(s) = f(s) + [g(s) + h(s)].$$

- Multiplication by a scalar is commutative, associative, and distributive:

$$k\,f(s) = f(s)\,k, \quad k[f(s) + g(s)] = kf(s) + kg(s),$$

$$(k_1 + k_2)f(s) = k_1 f(s) + k_2 f(s), \quad k_1[k_2 f(s)] = k_1 k_2 f(s).$$

- $\mathcal{H}$ is **spanned** by a set of basis functions φ_i, where for the purposes of this book the number of such basis functions (the range of i) can either be finite or denumerably infinite (like the positive integers). This means that every function in $\mathcal{H}$ can be represented by the linear form $f(s) = \sum_n a_n \varphi_n(s)$. This property is also known as **completeness**. We require that the basis functions be linearly independent, so that each function in the space will be a unique linear combination of the basis functions.
- For all functions $f(s)$ and $g(s)$ in $\mathcal{H}$, there exists a scalar product, denoted as $\langle f|g\rangle$, which evaluates to a finite real or complex numerical value (i.e., does not contain s) and which has the properties that

1. $\langle f|f\rangle \geq 0$, with the equality holding only if f is identically zero.[2] The quantity $\langle f|f\rangle^{1/2}$ is called the **norm** of f and is written $||f||$.
2. $\langle g|f\rangle^* = \langle f|g\rangle$, $\langle f|g+h\rangle = \langle f|g\rangle + \langle f|h\rangle$, and $\langle f|kg\rangle = k\langle f|g\rangle$.

Consequences of these properties are that $\langle f|k_1 g + k_2 h\rangle = k_1\langle f|g\rangle + k_2\langle f|h\rangle$, but $\langle kf|g\rangle = k^*\langle f|g\rangle$ and $\langle k_1 f + k_2 g|h\rangle = k_1^*\langle f|h\rangle + k_2^*\langle g|h\rangle$.

IN SUMMARY:

- Vector spaces that are
- closed under addition and
- multiplication by a scalar and
- which have a scalar product that exists for all pairs of its members
- are termed **Hilbert spaces**

Scalar Products

Now let's evaluate two scalar products:

$$\langle P_0|s^2\rangle = \int_{-1}^{1} P_0^*(s)s^2\,ds = \int_{-1}^{1}(1)(s^2)dx = \left[\frac{s^3}{3}\right]_{-1}^{1} = \frac{2}{3},$$

$$\langle P_0|P_2\rangle = \int_{-1}^{1}(1)\left[\frac{3}{2}s^2 - \frac{1}{2}\right]ds = \left[\frac{3}{2}\frac{s^3}{3} - \frac{1}{2}s\right]_{-1}^{1} = 0. \qquad (5.7)$$

Assignment: Find the other scalar products between elements of basis set S (Eq.(5.4) and S_1 (Eq.(5.5).

Orthogonal Expansions

With now a well-behaved scalar product in hand, we can make the definition that two functions f and g are **orthogonal** if $\langle f|g\rangle = 0$, which means that $\langle g|f\rangle$ will also vanish. An example of two functions that are orthogonal under the then-applicable definition of the scalar product are $P_0(s)$ and $P_2(s)$, where the scalar product is that defined in Eq. (5.6) and P_0, P_2 are the functions from Example 5.1.1; the orthogonality is shown by Eq. (5.7). We further define a function f as **normalized** if the scalar product $\langle f|f\rangle = 1$; this is the function-space equivalent of a unit vector. We will find that great convenience results if the basis functions for our function space are normalized and mutually orthogonal, corresponding to the description of a 2-D or three-dimensional (3-D) physical vector space based on orthogonal unit vectors. A set of functions that is both normalized and mutually orthogonal is called an **orthonormal** set. If a member f of an orthogonal set is not normalized, it can be made so without disturbing the orthogonality: we simply rescale it to $\overline{f} = f/\langle f|f\rangle^{1/2}$, so any orthogonal set can easily be made orthonormal if desired.

If our basis is orthonormal, the coefficients for the expansion of an arbitrary function in that basis take a simple form. We return to our 2-D example, with the assumption that the φ_i are orthonormal, and consider the result of taking the scalar product of $f(s)$, as given by Eq. (5.1), with $\varphi_1(s)$:

$$\langle\varphi_1|f\rangle = \langle\varphi_1|(a_1\varphi_1 + a_2\varphi_2)\rangle = a_1\langle\varphi_1|\varphi_1\rangle + a_2\langle\varphi_1|\varphi_2\rangle. \qquad (5.11)$$

The orthonormality of the φ now comes into play; the scalar product multiplying a_1 is unity, while that multiplying a_2 is zero, so we have the simple and useful result $\langle\varphi_1|f\rangle = a_1$. Thus, we have a rather mechanical means of identifying the components of f. The general result corresponding to Eq. (5.11) follows:

$$\text{If } \langle\varphi_i|\varphi_j\rangle = \delta_{ij} \quad\text{and}\quad f = \sum_{i=1}^{n} a_i\varphi_i, \quad\text{then}\quad a_i = \langle\varphi_i|f\rangle. \qquad (5.12)$$

Here the **Kronecker delta**, δ_{ij}, is unity if $i = j$ and zero otherwise. Looking once again at Eq. (5.11), we consider what happens if the φ_i are orthogonal but not normalized. Then instead of Eq. (5.12) we would have:

$$\text{If the } \varphi_i \text{ are orthogonal and } f = \sum_{i=1}^{n} a_i\varphi_i, \quad\text{then}\quad a_i = \frac{\langle\varphi_i|f\rangle}{\langle\varphi_i|\varphi_i\rangle}. \qquad (5.13)$$

This form of the expansion will be convenient when normalization of the basis introduces unpleasant factors.

Expansions and Scalar Products

If we have found the expansions of two functions, $f = \sum_\mu a_\mu \varphi_\mu$ and $g = \sum_\nu b_\nu \varphi_\nu$,

then their scalar product can be written $\langle f|g\rangle = \sum_{\mu\nu} a_\mu^* b_\nu \langle \varphi_\mu|\varphi_\nu\rangle$.

If the ' set is orthonormal, the above reduces to $\langle f|g\rangle = \sum_\mu a_\mu^* b_\mu$.

In the special case *g* = *f* , this reduces to $\langle f|f\rangle = \sum_\mu |a_\mu|^2$,

consistent with the requirement that $\langle f|f\rangle \geq 0$, with equality only if f is zero "almost everywhere."

If we regard the set of expansion coefficients a_μ as the elements of a column vector **a** representing f, with column vector **b** similarly representing g, Eqs. (5.18) and (5.19) correspond to the matrix equations

$$\langle f|g\rangle = \mathbf{a}^\dagger \mathbf{b}, \quad \langle f|f\rangle = \mathbf{a}^\dagger \mathbf{a}. \tag{5.20}$$

Note that by taking the adjoint of **a**, we both complex conjugate it and convert it into a row vector, so that the matrix products in Eq. (5.20) collapse to scalars, as required.

Schwarz Inequality

Any scalar product that meets the Hilbert space conditions will satisfy the **Schwarz inequality**, which can be stated as

$$|\langle f|g\rangle|^2 \leq \langle f|f\rangle\langle g|g\rangle. \tag{5.8}$$

Here there is equality only if f and g are proportional. In ordinary vector space, the equivalent result is, referring to Eq. (1.113),

$$(\mathbf{A}\cdot\mathbf{B})^2 = |\mathbf{A}|^2|\mathbf{B}|^2 \cos^2\theta \leq |\mathbf{A}|^2|\mathbf{B}|^2, \tag{5.9}$$

where θ is the angle between the directions of **A** and **B**. As observed previously, the equality only holds if **A** and **B** are collinear. If we also require **A** to be of unit length, we have the intuitively obvious result that the projection of **B** onto a noncollinear **A** direction will have a magnitude less than that of **B**. The Schwarz inequality extends this property to functions; their norms shrink on nontrivial projection.

Bessel's Inequality

A not too practical test for completeness is provided by **Bessel's inequality**, which states that if a function f has been expanded in an orthonormal basis as $\sum_n a_n\varphi_n$, then

$$\langle f|f\rangle \geq \sum_n |a_n|^2, \tag{5.21}$$

with the inequality occurring if the expansion of f is incomplete. The impracticality of this as a completeness test is that one needs to apply it for all f before using it to claim completeness of the space.

We establish Bessel's inequality by considering

$$I = \left\langle f - \sum_i a_i \varphi_i \middle| f - \sum_j a_j \varphi_j \right\rangle \geq 0, \qquad (5.22)$$

where $I = 0$ represents what is termed **convergence in the mean**, a criterion that permits the integrand to deviate from zero at isolated points. Expanding the scalar product, and eliminating terms that vanish because the φ are orthonormal, we arrive at Eq. (5.21), with equality only resulting if the expansion converges to f. We note in passing that convergence in the mean is a less stringent requirement than **uniform convergence**, but is adequate for almost all physical applications of basis-set expansions.

- To derive Bessel's inequality from Eq.(5.22)

5.1.7 Starting from $I = \left\langle f - \sum_i a_i \varphi_i \middle| f - \sum_j a_j \varphi_j \right\rangle \geq 0$,

derive Bessel's inequality, $\langle f|f\rangle \geq \sum_n |a_n|^2$.

Solution:

5.1.7. The φ_j are assumed to be orthonormal. Expanding I, we have

$$I = \langle f|f\rangle - \sum_i a_i^* \langle \varphi_i|f\rangle - \sum_i a_i \langle f|\varphi_i\rangle + \sum_{ij} a_i^* a_j \langle \varphi_i|\varphi_j\rangle \geq 0\,.$$

Using the relation $a_i = \langle \varphi_i|f\rangle$ and the orthonormality condition $\langle \varphi_i|\varphi_j\rangle = \delta_{ij}$,

$$I = \langle f|f\rangle - \sum_i a_i^* a_i - \sum_i a_i a_i^* + \sum_i a_i^* a_i = \langle f|f\rangle - \sum_i |a_i|^2 \geq 0\,.$$

Dirac Notation

Much of what we have discussed can be brought to a form that promotes clarity and suggests possibilities for additional analysis by using a notational device invented by P. A. M. Dirac. Dirac suggested that instead of just writing a function f, it be written enclosed in the right half of an angle-bracket pair, which he named a **ket**. Thus $f \to |f\rangle$, $\varphi_i \to |\varphi_i\rangle$, etc. Then he suggested that the complex conjugates of functions be enclosed in left half-brackets, which he named **bras**. An example of a bra is $\varphi_i^* \to \langle\varphi_i|$. Finally, he suggested that when the sequence (bra followed by ket = bra+ket ~ bracket) is encountered, the pair should be interpreted as a scalar product (with the dropping of one of the two adjacent vertical lines). As an initial example of the use of this notation, take Eq. (5.12), which we now write as

$$|f\rangle = \sum_j a_j |\varphi_j\rangle = \sum_j |\varphi_j\rangle\langle\varphi_j|f\rangle = \left(\sum_j |\varphi_j\rangle\langle\varphi_j|\right)|f\rangle. \qquad (5.31)$$

This notational rearrangement shows that we can view the expansion in the φ basis as the insertion of a set of basis members in a way which, in sum, has no effect. If the sum is over a complete set of φ_j, the ket-bra sum in Eq. (5.31) will have no net effect when inserted before any ket in the space, and therefore we can view the sum as a **resolution of the identity**. To emphasize this, we write

$$1 = \sum_j |\varphi_j\rangle\langle\varphi_j|. \qquad (5.32)$$

Many expressions involving expansions in orthonormal sets can be derived by the insertion of resolutions of the identity.

Dirac notation can also be applied to expressions involving vectors and matrices, where it illuminates the parallelism between physical vector spaces and the function spaces here under study. If $\mathbf{a}$ and $\mathbf{b}$ are column vectors and M is a matrix, then we can write $|\mathbf{b}\rangle$ as a synonym for $\mathbf{b}$, we can write $\langle\mathbf{a}|$ to mean $\mathbf{a}^\dagger$, and then $\langle\mathbf{a}|\mathbf{b}\rangle$ is interpreted as equivalent to $\mathbf{a}^\dagger\mathbf{b}$, which (when the vectors are real) is matrix notation for the (scalar) dot product $\mathbf{a}\cdot\mathbf{b}$. Other examples are expressions such as

$$\mathbf{a}=\mathsf{M}\mathbf{b}\leftrightarrow|\mathbf{a}\rangle=|\mathsf{M}\mathbf{b}\rangle=\mathsf{M}|\mathbf{b}\rangle \quad \text{or} \quad \mathbf{a}^\dagger\mathsf{M}\mathbf{b}=(\mathsf{M}^\dagger\mathbf{a})^\dagger\mathbf{b}\leftrightarrow\langle\mathbf{a}|\mathsf{M}\mathbf{b}\rangle=\langle\mathsf{M}^\dagger\mathbf{a}|\mathbf{b}\rangle.$$

Example: The matrix representation of unit vectors in ordinary space is given below.

(i) Find various scalar products from them :

Solution:

$$|i> = \begin{bmatrix}1\\0\\0\end{bmatrix}, |j> = \begin{bmatrix}0\\1\\0\end{bmatrix}, |k> = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

$$<i| = \begin{bmatrix}1\\0\\0\end{bmatrix}^\dagger = [\,1\;\;0\;\;0\,]\,, <j| = \begin{bmatrix}0\\1\\0\end{bmatrix}^\dagger = [\,0\;\;1\;\;0\,]\,, <k| = \begin{bmatrix}0\\0\\1\end{bmatrix}^\dagger = [\,0\;\;0\;\;1\,]$$

$$<i||i> = \begin{bmatrix}1\\0\\0\end{bmatrix}^\dagger \begin{bmatrix}1\\0\\0\end{bmatrix} = [\,1\;\;0\;\;0\,]\begin{bmatrix}1\\0\\0\end{bmatrix} = 1 \quad \text{and}$$

$$|i><i| = \begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}^\dagger = \begin{bmatrix}1\\0\\0\end{bmatrix}[\,1\;\;0\;\;0\,] = \begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}$$

Similarly find $<j||j>$, $|j><j|$ and $<k||k>$, $|k><k|$

(ii) Also prove the resolution of identity: $\sum_n |n><n| = I$

Solution: Consider

$$\sum_n |n><n| = |i><i| + |j><j| + |k><k|$$
$$= \begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix} + \begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix} + \begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$$
$$= \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} = I$$

Problems:

5.1.8 Expand the function $\sin\pi x$ in a series of functions φ_i that are orthogonal (but not normalized) on the range $0\le x\le 1$ when the scalar product has definition

$$\langle f|g\rangle = \int_0^1 f^*(x)g(x)\,dx.$$

Keep the first four terms of the expansion. The first four φ_i are:

$$\varphi_0 = 1, \quad \varphi_1 = 2x-1, \quad \varphi_2 = 6x^2-6x+1, \quad \varphi_3 = 20x^3-30x^2+12x-1.$$

Note. The integrals that are needed are the subject of Example 1.10.5.

Answer:

$$\sin \pi x = \frac{2/\pi}{1}\varphi_0 + \frac{2/\pi - 24/\pi^3}{1/5}\varphi_2 + \cdots = 0.6366 - 0.6871(6x^2 - 6x + 1) + \cdots .$$

This series converges fairly rapidly. See Fig. 5.1.8.

5.1.9 Expand the function e^{-x} in Laguerre polynomials $L_n(x)$, which are orthonormal on the range $0 \le x < \infty$ with scalar product

$$\langle f|g\rangle = \int_0^\infty f^*(x)g(x)e^{-x}\,dx.$$

Keep the first four terms of the expansion. The first four $L_n(x)$ are

$$L_0 = 1, \quad L_1 = 1 - x, \quad L_2 = \frac{2 - 4x + x^2}{2}, \quad L_3 = \frac{6 - 18x + 9x^2 - x^3}{6}.$$

Solution:

9. $e^{-x} = a_0 L_0(x) + a_1 L_1(x) + a_2 L_2(x) + a_3 L_3(x) + \cdots ,$

$$a_i = \int_0^\infty L_i(x) e^{-2x}\,dx .$$

By integration we find $a_0 = 1/2$, $a_1 = 1/4$, $a_2 = 1/8$, $a_3 = 1/16$. Thus,

$$e^{-x} = \frac{1}{2}(1) + \frac{1}{4}(1 - x) + \frac{1}{8}\frac{2 - 4x + x^2}{2} + \frac{1}{16}\frac{6 - 18x + 9x^2 - x^3}{6} + \cdots .$$

This expansion when terminated after L_3 fails badly beyond about $x = 3$. See Fig. 5.1.9.

---

10 The explicit form of a function f is not known, but the coefficients a_n of its expansion in the orthonormal set φ_n are available. Assuming that the φ_n and the members of another orthonormal set, χ_n, are available, use Dirac notation to obtain a formula for the coefficients for the expansion of f in the χ_n set.

Solution:

10. The forms $\sum_i |\varphi_i\rangle\langle\varphi_i|$ and $\sum_j |\chi_j\rangle\langle\chi_j|$ are resolutions of the identity. Therefore

$$|f\rangle = \sum_{ij} |\chi_j\rangle\langle\chi_j|\varphi_i\rangle\langle\varphi_i|f\rangle .$$

The coefficients of f in the φ basis are $a_i = \langle\varphi_i|f\rangle$, so the above equation is equivalent to

$$f = \sum_j b_j \chi_j, \quad \text{with} \quad b_j = \sum_i \langle\chi_j|\varphi_i\rangle a_i .$$

.11 Using conventional vector notation, evaluate $\sum_j |\hat{\mathbf{e}}_j\rangle\langle\hat{\mathbf{e}}_j|\mathbf{a}\rangle$, where $\mathbf{a}$ is an arbitrary vector in the space spanned by the $\hat{\mathbf{e}}_j$.

Solution

11. We assume the unit vectors are orthogonal. Then,

$$\sum_j |\hat{\mathbf{e}}_j\rangle\langle\hat{\mathbf{e}}_j|\mathbf{a}\rangle = \sum_j (\hat{\mathbf{e}}_j \cdot \mathbf{a})\,\hat{\mathbf{e}}_j\,.$$

:

This expression is a component decomposition of $\mathbf{a}$.

.12 Letting $\mathbf{a} = a_1\hat{\mathbf{e}}_1 + a_2\hat{\mathbf{e}}_2$ and $\mathbf{b} = b_1\hat{\mathbf{e}}_1 + b_2\hat{\mathbf{e}}_2$ be vectors in $\mathbb{R}^2$, for what values of k, if any, is

$$\langle\mathbf{a}|\mathbf{b}\rangle = a_1b_1 - a_1b_2 - a_2b_1 + ka_2b_2$$

a valid definition of a scalar product?

Solution:

12. The scalar product $\langle\mathbf{a}|\mathbf{a}\rangle$ must be positive for every nonzero vector in the space. If we write $\langle\mathbf{a}|\mathbf{a}\rangle$ in the form $(a_1 - a_2)^2 + (k-1)a_2^2$, this condition will be violated for some nonzero $\mathbf{a}$ unless $k > 1$.

5.2 GRAM-SCHMIDT ORTHOGONALIZATION

Crucial to carrying out the expansions and transformations under discussion is the availability of useful orthonormal sets of functions. We therefore proceed to the description of a process whereby a set of functions that is neither orthogonal or normalized can be used to construct an orthonormal set that spans the same function space. There are many ways to accomplish this task. We present here the method called the **Gram-Schmidt** orthogonalization process.

The Gram-Schmidt process assumes the availability of a set of functions χ_μ and an appropriately defined scalar product $\langle f|g\rangle$. We orthonormalize **sequentially** to form the orthonormal functions φ_ν, meaning we make the first orthonormal function, φ_0, from χ_0, the next, φ_1, from χ_0 and χ_1, etc. If, for example, the χ_μ are powers x^μ, the orthonormal function φ_ν will be a polynomial of degree ν in x. Because the Gram-Schmidt process is often applied to powers, we have chosen to number both the χ and the φ sets starting from zero (rather than 1).

Thus, our first orthonormal function will simply be a normalized version of χ_0. Specifically,

$$\varphi_0 = \frac{\chi_0}{\langle\chi_0|\chi_0\rangle^{1/2}}. \tag{5.33}$$

To check that Eq. (5.33) is correct, we form

$$\langle\varphi_0|\varphi_0\rangle = \left\langle \frac{\chi_0}{\langle\chi_0|\chi_0\rangle^{1/2}} \middle| \frac{\chi_0}{\langle\chi_0|\chi_0\rangle^{1/2}} \right\rangle = 1.$$

Next, starting from φ_0 and χ_1, we form a function that is orthogonal to φ_0. We use φ_0 rather than χ_0 to be consistent with what we will do in later steps of the process. Thus, we write

$$\psi_1 = \chi_1 - a_{1,0}\varphi_0. \tag{5.34}$$

What we are doing here is the removal from χ_1 of its projection onto φ_0, leaving a remainder that will be orthogonal to φ_0. Remembering that φ_0 is normalized (of "unit length"), that projection is identified as $\langle\varphi_0|\chi_1\rangle\varphi_0$, so that

$$a_{1,0} = \langle\varphi_0|\chi_1\rangle. \tag{5.35}$$

In case Eq. (5.35) is not intuitively obvious, we can confirm it by writing the requirement that ψ_1 be orthogonal to φ_0:

$$\langle\varphi_0|\psi_1\rangle = \Big\langle\varphi_0\Big|\,\big(\chi_1 - a_{1,0}\varphi_0\big)\Big\rangle = \langle\varphi_0|\chi_1\rangle - a_{1,0}\langle\varphi_0|\varphi_0\rangle = 0,$$

which, because φ_0 is normalized, reduces to Eq. (5.35). The function ψ_1 is not in general normalized. To normalize it and thereby obtain φ_1, we form

$$\varphi_1 = \frac{\psi_1}{\langle\psi_1|\psi_1\rangle^{1/2}}. \tag{5.36}$$

To continue further, we need to make, from φ_0, φ_1, and χ_2, a function that is orthogonal to both φ_0 and φ_1. It will have the form

$$\psi_2 = \chi_2 - a_{0,2}\varphi_0 - a_{1,2}\varphi_1. \tag{5.37}$$

The last two terms of Eq. (5.37), respectively, remove from χ_2 its projections on φ_0 and φ_1; these projections are independent because φ_0 and φ_1 are orthogonal. Thus, either from our knowledge of projections or by setting to zero the scalar products $\langle\varphi_i|\psi_2\rangle$ ($i = 0$ and 1), we establish

$$a_{0,2} = \langle\varphi_0|\chi_2\rangle, \quad a_{1,2} = \langle\varphi_1|\chi_2\rangle. \tag{5.38}$$

Finally, we make $\varphi_2 = \psi_2/\langle\psi_2|\psi_2\rangle^{1/2}$.

The generalization for which the above is the first few terms is that, given the prior formation of φ_i, $i = 0, \dots, n-1$, the orthonormal function φ_n is obtained from χ_n by the following two steps:

$$\psi_n = \chi_n - \sum_{\mu=0}^{n-1}\langle\varphi_\mu|\chi_n\rangle\varphi_\mu,$$

$$\varphi_n = \frac{\psi_n}{\langle\psi_n|\psi_n\rangle^{1/2}}. \tag{5.39}$$

Reviewing the above process, we note that different results would have been obtained if we used the same set of χ_i, but simply took them in a different order. For example, if we had started with χ_3, one of our orthonormal functions would have been a multiple of χ_3, while the set we constructed yielded φ_3 as a linear combination of χ_μ, $\mu = 0, 1, 2, 3$.

Example 5.2.1 Legendre Polynomials

Let us form an orthonormal set, taking the χ_μ as x^μ, and making the definition

$$\langle f|g\rangle = \int_{-1}^{1} f^*(x)g(x)dx. \tag{5.40}$$

Solution:

This scalar product definition will cause the members of our set to be orthogonal, with unit weight, on the range $(-1, 1)$. Moreover, since the χ_μ are real, the complex conjugate asterisk has no operational significance here.

The first orthonormal function, φ_0, is

$$\varphi_0(x) = \frac{1}{\langle 1|1\rangle^{1/2}} = \frac{1}{\left[\int_{-1}^{1} dx\right]^{1/2}} = \frac{1}{\sqrt{2}}.$$

To obtain φ_1, we first obtain ψ_1 by evaluating

$$\psi_1(x) = x - \langle\varphi_0|x\rangle\varphi_0(x) = x,$$

where the scalar product vanishes because φ_0 is an even function of x, whereas x is odd, and the range of integration is even. We then find

$$\varphi_1(x) = \frac{x}{\left[\int_{-1}^{1} x^2 dx\right]^{1/2}} = \sqrt{\frac{3}{2}}\,x.$$

The next step is less trivial. We form

$$\psi_2(x) = x^2 - \langle\varphi_0|x^2\rangle\varphi_0(x) - \langle\varphi_1|x^2\rangle\varphi_1(x) = x^2 - \left\langle\frac{1}{\sqrt{2}}\middle|x^2\right\rangle\left(\frac{1}{\sqrt{2}}\right) = x^2 - \frac{1}{3},$$

where we have used symmetry to set $\langle\varphi_1|x^2\rangle$ to zero and evaluated the scalar product

$$\left\langle\frac{1}{\sqrt{2}}\middle|x^2\right\rangle = \frac{1}{\sqrt{2}}\int_{-1}^{1} x^2 dx = \frac{\sqrt{2}}{3}.$$

Then,

$$\varphi_2(x) = \frac{x^2 - \frac{1}{3}}{\left[\int_{-1}^{1} (x^2 - \frac{1}{3})^2 dx\right]^{1/2}} = \sqrt{\frac{5}{2}}\left(\frac{3}{2}x^2 - \frac{1}{2}\right).$$

Continuation to one more orthonormal function yields

$$\varphi_3(x) = \sqrt{\frac{7}{2}}\left(\frac{5}{2}x^3 - \frac{3}{2}x\right).$$

Reference to Chapter 15 will show that

$$\varphi_n(x) = \sqrt{\frac{2n+1}{2}}\,P_n(x), \tag{5.41}$$

where $P_n(x)$ is the nth degree Legendre polynomial. Our Gram-Schmidt process provides a possible but very cumbersome method of generating the Legendre polynomials; other, more efficient approaches exist. ■

Orthonormalizing Physical Vectors

The Gram-Schmidt process also works for ordinary vectors that are simply given by their components, it being understood that the scalar product is just the ordinary dot product.

5.2.8 Form a set of three orthonormal vectors by the Gram-Schmidt process using these input vectors in the order given:

$$\mathbf{c}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \mathbf{c}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

Solution:

5.2.8. Let the orthonormalized vectors be denoted $\mathbf{b}_i$. First, Make $\mathbf{b}_1$ a normalized version of $\mathbf{c}_1$: $\mathbf{b}_1 = \mathbf{c}_1/\sqrt{3}$. Then obtain $\overline{b}_2$ (denoting $\mathbf{b}_2$ before normalization) as

$$\overline{\mathbf{b}}_2 = \mathbf{c}_2 - (\mathbf{b}_1 \cdot \mathbf{c}_2)\mathbf{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} - (4\sqrt{3}) \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} = \begin{pmatrix} -1/3 \\ -1/3 \\ 2/3 \end{pmatrix}.$$

Normalizing, $\mathbf{b}_2 = \sqrt{3/2}\,\overline{\mathbf{b}}_2$. Finally, form

$$\overline{\mathbf{b}}_3 = \mathbf{c}_3 - (\mathbf{b}_1 \cdot \mathbf{c}_3)\mathbf{b}_1 - (\mathbf{b}_2 \cdot \mathbf{c}_3)\mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} - (\sqrt{3}) \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$$

$$- (\sqrt{3}/2) \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix} = \begin{pmatrix} 1/2 \\ -1/2 \\ 0 \end{pmatrix}.$$

Normalizing, $\mathbf{b}_3 = \sqrt{2}\,\overline{\mathbf{b}}_3$. Collecting our answers, the orthonormal vectors are

$$\mathbf{b}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix}, \quad \mathbf{b}_3 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}.$$

xxxxxxxxxxxxxxxxxxxx