Inner Product Spaces and Orthogonal Functions

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts at Lowell
Lowell, MA 01854, USA

May 1, 2008

1 Background

We begin by recalling the solution of the vibrating string problem and Sturm-Liouville problems.

1.1 The Vibrating String

When we solve the problem of the vibrating string using the technique of separation of variables, the differential equation involving the space variable x, and assuming constant mass density, is

$$y''(x) + \frac{\omega^2}{c^2} y(x) = 0, \tag{1.1}$$

which we can write as an eigenvalue problem

$$y''(x) + \lambda y(x) = 0. \tag{1.2}$$

The solutions to Equation (1.1) are

$$y(x) = \alpha \sin\Big(\frac{\omega}{c} x\Big).$$

In the vibrating string problem, the string is fixed at both ends, $x = 0$ and $x = L$, so that

$$\phi(0, t) = \phi(L, t) = 0,$$

for all t. Therefore, we must have $y(0) = y(L) = 0$, so that the *eigenfunction solution* that corresponds to the eigenvalue $\lambda_m = \left(\frac{\pi m}{L}\right)^2$ must have the form

$$y(x) = A_m \sin\Big(\frac{\omega_m}{c} x\Big) = A_m \sin\Big(\frac{\pi m}{L} x\Big),$$

where $\omega_m = \frac{\pi c m}{L}$, for any positive integer m. Therefore, the boundary conditions limit the choices for the separation constant ω.

We then discover that the eigenfunction solutions corresponding to different λ are *orthogonal*, in the sense that

$$\int_0^L \sin\left(\frac{\pi m}{L}x\right)\sin\left(\frac{\pi n}{L}x\right)dx = 0,$$

for $m \neq n$.

1.2 The Sturm-Liouville Problem

The general form for the Sturm-Liouville Problem is

$$\frac{d}{dx}\Big(p(x)y'(x)\Big) + \lambda w(x)y(x) = 0. \tag{1.3}$$

As with the one-dimensional wave equation, boundary conditions, such as $y(a) = y(b) = 0$, where $a = -\infty$ and $b = +\infty$ are allowed, restrict the possible eigenvalues λ to an increasing sequence of positive numbers λ_m. The corresponding eigenfunctions $y_m(x)$ will be $w(x)$-orthogonal, meaning that

$$0 = \int_a^b y_m(x)y_n(x)w(x)dx,$$

for $m \neq n$. For various choices of $w(x)$ and $p(x)$ and various choices of a and b, we obtain several famous sets of "orthogonal" functions.

Well known examples of Sturm-Liouville problems include

- **Legendre:**
$$\frac{d}{dx}\Big((1-x^2)\frac{dy}{dx}\Big) + \lambda y = 0;$$
- **Chebyshev:**
$$\frac{d}{dx}\Big(\sqrt{1-x^2}\frac{dy}{dx}\Big) + \lambda(1-x^2)^{-1/2}y = 0;$$
- **Hermite:**
$$\frac{d}{dx}\Big(e^{-x^2}\frac{dy}{dx}\Big) + \lambda e^{-x^2}y = 0;$$

  and
- **Laguerre:**
$$\frac{d}{dx}\Big(xe^{-x}\frac{dy}{dx}\Big) + \lambda e^{-x}y = 0.$$

Each of these examples involves an inner product space and an orthogonal basis for that space.

2 The Complex Vector Dot Product

An *inner product* is a generalization of the notion of the dot product between two complex vectors.

2.1 The Two-Dimensional Case

Let $\mathbf{u} = (a, b)$ and $\mathbf{v} = (c, d)$ be two vectors in two-dimensional space. Let $\mathbf{u}$ make the angle $\alpha > 0$ with the positive x-axis and $\mathbf{v}$ the angle $\beta > 0$. Let $||\mathbf{u}|| = \sqrt{a^2 + b^2}$ denote the length of the vector $\mathbf{u}$. Then $a = ||\mathbf{u}|| \cos\alpha$, $b = ||\mathbf{u}|| \sin\alpha$, $c = ||\mathbf{v}|| \cos\beta$ and $d = ||\mathbf{v}|| \sin\beta$. So $\mathbf{u} \cdot \mathbf{v} = ac + bd = ||\mathbf{u}|| ||\mathbf{v}|| (\cos\alpha\cos\beta + \sin\alpha\sin\beta = ||\mathbf{u}|| \, ||\mathbf{v}|| \cos(\alpha - \beta)$. Therefore, we have

$$\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}|| \, ||\mathbf{v}|| \cos\theta, \tag{2.1}$$

where $\theta = \alpha - \beta$ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Cauchy's inequality is

$$|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| \, ||\mathbf{v}||,$$

with equality if and only if $\mathbf{u}$ and $\mathbf{v}$ are parallel. From Equation (2.1) we know that the dot product $\mathbf{u} \cdot \mathbf{v}$ is zero if and only if the angle between these two vectors is a right angle; we say then that $\mathbf{u}$ and $\mathbf{v}$ are mutually *orthogonal*.

Cauchy's inequality extends to complex vectors $\mathbf{u}$ and $\mathbf{v}$:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{n=1}^{N} u_n \overline{v_n}, \tag{2.2}$$

and Cauchy's Inequality still holds.

Proof of Cauchy's inequality: To prove Cauchy's inequality for the complex vector dot product, we write $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u} \cdot \mathbf{v}| e^{i\theta}$. Let t be a real variable and consider

$$0 \leq ||e^{-i\theta}\mathbf{u} - t\mathbf{v}||^2 = (e^{-i\theta}\mathbf{u} - t\mathbf{v}) \cdot (e^{-i\theta}\mathbf{u} - t\mathbf{v})$$

$$= ||\mathbf{u}||^2 - t[(e^{-i\theta}\mathbf{u}) \cdot \mathbf{v} + \mathbf{v} \cdot (e^{-i\theta}\mathbf{u})] + t^2||\mathbf{v}||^2$$

$$= ||\mathbf{u}||^2 - t[(e^{-i\theta}\mathbf{u}) \cdot \mathbf{v} + \overline{(e^{-i\theta}\mathbf{u}) \cdot \mathbf{v}}] + t^2||\mathbf{v}||^2$$

$$= ||\mathbf{u}||^2 - 2Re(te^{-i\theta}(\mathbf{u} \cdot \mathbf{v})) + t^2||\mathbf{v}||^2$$

$$= ||\mathbf{u}||^2 - 2Re(t|\mathbf{u} \cdot \mathbf{v}|) + t^2||\mathbf{v}||^2 = ||\mathbf{u}||^2 - 2t|\mathbf{u} \cdot \mathbf{v}| + t^2||\mathbf{v}||^2.$$

This is a nonnegative quadratic polynomial in the variable t, so it cannot have two distinct real roots. Therefore, the discriminant $4|\mathbf{u} \cdot \mathbf{v}|^2 - 4||\mathbf{v}||^2||\mathbf{u}||^2$ must be non-positive; that is, $|\mathbf{u} \cdot \mathbf{v}|^2 \leq ||\mathbf{u}||^2||\mathbf{v}||^2$. This is Cauchy's inequality. ∎

A careful examination of the proof just presented shows that we did not explicitly use the definition of the complex vector dot product, but only some of its properties. This suggested to mathematicians the possibility of abstracting these properties and using them to define a more general concept, an *inner product*, between objects more general than complex vectors, such as infinite sequences, random variables, and matrices. Such an inner product can then be used to define the *norm* of these objects and thereby a distance between such objects. Once we have an inner product defined, we also have available the notions of orthogonality and best approximation.

2.2 Orthogonality

Consider the problem of writing the two-dimensional real vector $(3,-2)$ as a linear combination of the vectors $(1,1)$ and $(1,-1)$; that is, we want to find constants a and b so that $(3,-2)=a(1,1)+b(1,-1)$. One way to do this, of course, is to compare the components: $3=a+b$ and $-2=a-b$; we can then solve this simple system for the a and b. In higher dimensions this way of doing it becomes harder, however. A second way is to make use of the dot product and orthogonality.

The dot product of two vectors (x,y) and (w,z) in R^2 is $(x,y)\cdot(w,z)=xw+yz$. If the dot product is zero then the vectors are said to be *orthogonal*; the two vectors $(1,1)$ and $(1,-1)$ are orthogonal. We take the dot product of both sides of $(3,-2)=a(1,1)+b(1,-1)$ with $(1,1)$ to get

$$1=(3,-2)\cdot(1,1)=a(1,1)\cdot(1,1)+b(1,-1)\cdot(1,1)=a(1,1)\cdot(1,1)+0=2a,$$

so we see that $a=\frac{1}{2}$. Similarly, taking the dot product of both sides with $(1,-1)$ gives

$$5=(3,-2)\cdot(1,-1)=a(1,1)\cdot(1,-1)+b(1,-1)\cdot(1,-1)=2b,$$

so $b=\frac{5}{2}$. Therefore, $(3,-2)=\frac{1}{2}(1,1)+\frac{5}{2}(1,-1)$. The beauty of this approach is that it does not get much harder as we go to higher dimensions.

Since the cosine of the angle θ between vectors $\mathbf{u}$ and $\mathbf{v}$ is

$$\cos\theta=\mathbf{u}\cdot\mathbf{v}/||\mathbf{u}||\,||\mathbf{v}||,$$

where $||\mathbf{u}||^2=\mathbf{u}\cdot\mathbf{u}$, the projection of vector $\mathbf{v}$ on to the line through the origin parallel to $\mathbf{u}$ is

$$\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})=\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}.$$

Therefore, the vector $\mathbf{v}$ can be written as

$$\mathbf{v}=\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})+(\mathbf{v}-\mathrm{Proj}_{\mathbf{u}}(\mathbf{v})),$$

where the first term on the right is parallel to $\mathbf{u}$ and the second one is orthogonal to $\mathbf{u}$.

How do we find vectors that are mutually orthogonal? Suppose we begin with $(1,1)$. Take a second vector, say $(1,2)$, that is not parallel to $(1,1)$ and write it as we did $\mathbf{v}$ earlier, that is, as a sum of two vectors, one parallel to $(1,1)$ and the second orthogonal to $(1,1)$. The projection of $(1,2)$ onto the line parallel to $(1,1)$ passing through the origin is

$$\frac{(1,1)\cdot(1,2)}{(1,1)\cdot(1,1)}(1,1) = \frac{3}{2}(1,1) = (\frac{3}{2},\frac{3}{2})$$

so

$$(1,2) = (\frac{3}{2},\frac{3}{2}) + ((1,2) - (\frac{3}{2},\frac{3}{2})) = (\frac{3}{2},\frac{3}{2}) + (-\frac{1}{2},\frac{1}{2}).$$

The vectors $(-\frac{1}{2},\frac{1}{2}) = -\frac{1}{2}(1,-1)$ and, therefore, $(1,-1)$ are then orthogonal to $(1,1)$. This approach is the basis for the *Gram-Schmidt* method for constructing a set of mutually orthogonal vectors.

3 Generalizing the Dot Product: Inner Products

The proof of Cauchy's Inequality rests not on the actual definition of the complex vector dot product, but rather on four of its most basic properties. We use these properties to extend the concept of the complex vector dot product to that of *inner product*. Later in this chapter we shall give several examples of inner products, applied to a variety of mathematical objects, including infinite sequences, functions, random variables, and matrices. For now, let us denote our mathematical objects by $\mathbf{u}$ and $\mathbf{v}$ and the inner product between them as $\langle \mathbf{u}, \mathbf{v} \rangle$. The objects will then be said to be members of an *inner-product space*. We are interested in inner products because they provide a notion of orthogonality, which is fundamental to best approximation and optimal estimation.

3.1 Defining an Inner Product and Norm

The four basic properties that will serve to define an inner product are:

- **1:** $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, with equality if and only if $\mathbf{u} = \mathbf{0}$;
- **2:** $\langle \mathbf{v}, \mathbf{u} \rangle = \overline{\langle \mathbf{u}, \mathbf{v} \rangle}$;
- **3:** $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$;

- **4:** $\langle c\mathbf{u}, \mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$ for any complex number c.

The inner product is the basic ingredient in Hilbert space theory. Using the inner product, we define the *norm* of $\mathbf{u}$ to be

$$||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$$

and the distance between $\mathbf{u}$ and $\mathbf{v}$ to be $||\mathbf{u} - \mathbf{v}||$.

The Cauchy-Schwarz inequality: Because these four properties were all we needed to prove the Cauchy inequality for the complex vector dot product, we obtain the same inequality whenever we have an inner product. This more general inequality is the Cauchy-Schwarz inequality:

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}\sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$$

or

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq ||\mathbf{u}||\,||\mathbf{v}||,$$

with equality if and only if there is a scalar c such that $\mathbf{v} = c\mathbf{u}$. We say that the vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. We turn now to some examples.

3.2 Some Examples of Inner Products

Here are several examples of inner products.

- **Inner product of infinite sequences:** Let $\mathbf{u} = \{u_n\}$ and $\mathbf{v} = \{v_n\}$ be infinite sequences of complex numbers. The inner product is then

  $$\langle \mathbf{u}, \mathbf{v}\rangle = \sum u_n \overline{v_n},$$

  and

  $$||\mathbf{u}|| = \sqrt{\sum |u_n|^2}.$$

  The sums are assumed to be finite; the index of summation n is singly or doubly infinite, depending on the context. The Cauchy-Schwarz inequality says that

  $$|\sum u_n \overline{v_n}| \leq \sqrt{\sum |u_n|^2}\sqrt{\sum |v_n|^2}.$$

- **Inner product of functions:** Now suppose that $\mathbf{u} = f(x)$ and $\mathbf{v} = g(x)$. Then,
$$\langle \mathbf{u}, \mathbf{v} \rangle \ = \int f(x)\overline{g(x)}dx$$
and
$$||\mathbf{u}|| = \sqrt{\int |f(x)|^2 dx}.$$
The integrals are assumed to be finite; the limits of integration depend on the support of the functions involved. The Cauchy-Schwarz inequality now says that
$$|\int f(x)\overline{g(x)}dx| \leq \sqrt{\int |f(x)|^2 dx}\sqrt{\int |g(x)|^2 dx}.$$

- **Inner product of random variables:** Now suppose that $\mathbf{u} = X$ and $\mathbf{v} = Y$ are random variables. Then,
$$\langle \mathbf{u}, \mathbf{v} \rangle \ = E(X\overline{Y})$$
and
$$||\mathbf{u}|| = \sqrt{E(|X|^2)},$$
which is the standard deviation of X if the mean of X is zero. The expected values are assumed to be finite. The Cauchy-Schwarz inequality now says that
$$|E(X\overline{Y})| \leq \sqrt{E(|X|^2)}\sqrt{E(|Y|^2)}.$$
If $E(X) = 0$ and $E(Y) = 0$, the random variables X and Y are orthogonal if and only if they are *uncorrelated*.

- **Inner product of complex matrices:** Now suppose that $\mathbf{u} = A$ and $\mathbf{v} = B$ are complex matrices. Then,
$$\langle \mathbf{u}, \mathbf{v} \rangle \ = \mathrm{trace}(B^\dagger A)$$
and
$$||\mathbf{u}|| = \sqrt{\mathrm{trace}(A^\dagger A)},$$
where the trace of a square matrix is the sum of the entries on the main diagonal. As we shall see later, this inner product is simply the complex vector dot product of the vectorized versions of the matrices involved. The Cauchy-Schwarz inequality now says that
$$|\mathrm{trace}(B^\dagger A)| \leq \sqrt{\mathrm{trace}(A^\dagger A)}\sqrt{\mathrm{trace}(B^\dagger B)}.$$

- **Weighted inner product of complex vectors:** Let $\mathbf{u}$ and $\mathbf{v}$ be complex vectors and let Q be a Hermitian positive-definite matrix; that is, $Q^\dagger = Q$ and $\mathbf{u}^\dagger Q\mathbf{u} > 0$ for all nonzero vectors $\mathbf{u}$. The inner product is then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^\dagger Q\mathbf{u}$$

and

$$||\mathbf{u}|| = \sqrt{\mathbf{u}^\dagger Q\mathbf{u}}.$$

We know from the eigenvector decomposition of Q that $Q = C^\dagger C$ for some matrix C. Therefore, the inner product is simply the complex vector dot product of the vectors $C\mathbf{u}$ and $C\mathbf{v}$. The Cauchy-Schwarz inequality says that

$$|\mathbf{v}^\dagger Q\mathbf{u}| \leq \sqrt{\mathbf{u}^\dagger Q\mathbf{u}}\sqrt{\mathbf{v}^\dagger Q\mathbf{v}}.$$

- **Weighted inner product of functions:** Now suppose that $\mathbf{u} = f(x)$ and $\mathbf{v} = g(x)$ and $w(x) > 0$. Then define

$$\langle \mathbf{u}, \mathbf{v} \rangle = \int f(x)\overline{g(x)}w(x)dx$$

and

$$||\mathbf{u}|| = \sqrt{\int |f(x)|^2 w(x)dx}.$$

The integrals are assumed to be finite; the limits of integration depend on the support of the functions involved. This inner product is simply the inner product of the functions $f(x)\sqrt{w(x)}$ and $g(x)\sqrt{w(x)}$. The Cauchy-Schwarz inequality now says that

$$|\int f(x)\overline{g(x)}w(x)dx| \leq \sqrt{\int |f(x)|^2 w(x)dx}\sqrt{\int |g(x)|^2 w(x)dx}.$$

Once we have an inner product defined, we can speak about orthogonality and best approximation. Important in that regard is the orthogonality principle.

4 Best Approximation and the Orthogonality Principle

Imagine that you are standing and looking down at the floor. The point B on the floor that is closest to N, the tip of your nose, is the unique point on the floor such that the vector from B to any other point A on the floor is perpendicular to the vector from N to B; that is, $\langle BN, BA\rangle = 0$. This is a simple illustration of the *orthogonality principle*. Whenever we have an inner product defined we can speak of orthogonality and apply the orthogonality principle to find best approximations. For notational simplicity, we shall consider only real inner product spaces.

4.1 Best Approximation

Let $\mathbf{u}$ and $\mathbf{v}^1, ..., \mathbf{v}^N$ be members of a real inner-product space. For all choices of scalars $a_1, ..., a_N$, we can compute the distance from $\mathbf{u}$ to the member $a_1\mathbf{v}^1 + ...a_N\mathbf{v}^N$. Then, we minimize this distance over all choices of the scalars; let $b_1, ..., b_N$ be this best choice.

The distance squared from $\mathbf{u}$ to $a_1\mathbf{v}^1 + ...a_N\mathbf{v}^N$ is

$$||\mathbf{u} - (a_1\mathbf{v}^1 + ...a_N\mathbf{v}^N)||^2 = \langle \mathbf{u} - (a_1\mathbf{v}^1 + ...a_N\mathbf{v}^N), \mathbf{u} - (a_1\mathbf{v}^1 + ...a_N\mathbf{v}^N)\rangle,$$

$$= ||\mathbf{u}||^2 - 2\langle \mathbf{u}, \sum_{n=1}^N a_n\mathbf{v}^n\rangle + \sum_{n=1}^N\sum_{m=1}^N a_n a_m \langle \mathbf{v}^n\mathbf{v}^m\rangle.$$

Setting the partial derivative with respect to a_n equal to zero, we have

$$\langle \mathbf{u}, \mathbf{v}^n\rangle = \sum_{m=1}^N a_m\langle \mathbf{v}^m\mathbf{v}^n\rangle.$$

With $\mathbf{a} = (a_1, ..., a_N)^T$,

$$\mathbf{d} = (\langle \mathbf{u}, \mathbf{v}^1\rangle, ..., \langle \mathbf{u}, \mathbf{v}^N\rangle)^T$$

and V the matrix with entries

$$V_{mn} = \langle \mathbf{v}^m\mathbf{v}^n\rangle,$$

we find that we must solve the system of equations $V\mathbf{a} = \mathbf{d}$. When the vectors $\mathbf{v}^n$ are mutually orthogonal and each has norm equal to one, then $V = I$, the identity matrix, and the desired vector $\mathbf{a}$ is simply $\mathbf{d}$.

4.2 The Orthogonality Principle

The *orthogonality principle* provides another way to view the calculation of the best approximation: let the best approximation of $\mathbf{u}$ be the vector

$$b_1\mathbf{v}^1 + ...b_N\mathbf{v}^N.$$

Then

$$\langle \mathbf{u} - b_1\mathbf{v}^1 + ...b_N\mathbf{v}^N, \mathbf{v}^n\rangle = 0,$$

for $n = 1, 2, ..., N$. This leads directly to the system of equations

$$\mathbf{d} = V\mathbf{b},$$

which, as we just saw, provides the optimal coefficients.

To see why the orthogonality principle is valid, fix a value of n and consider the problem of minimizing the distance

$$||\mathbf{u} - (b_1\mathbf{v}^1 + ...b_N\mathbf{v}^N + \alpha\mathbf{v}^n)||$$

as a function of α. Writing the norm squared in terms of the inner product, expanding the terms, and differentiating with respect to α, we find that the minimum occurs when

$$\alpha = \langle \mathbf{u} - b_1\mathbf{v}^1 + ...b_N\mathbf{v}^N, \mathbf{v}^n\rangle.$$

But we already know that the minimum occurs when $\alpha = 0$. This completes the proof of the orthogonality principle.

5 Gram-Schmidt Orthogonalization

We have seen that the best approximation is easily calculated if the vectors $\mathbf{v}^n$ are mutually orthogonal. But how to we get such a mutually orthogonal set, in general? The Gram-Schmidt Orthogonalization Method is one way to proceed.

Let $\{\mathbf{v}^1, ..., \mathbf{v}^N\}$ be a linearly independent set of vectors in the space R^M, where $N \leq M$. The Gram-Schmidt method uses the $\mathbf{v}^n$ to create an orthogonal basis $\{\mathbf{u}^1, ..., \mathbf{u}^N\}$ for the span of the $\mathbf{v}^n$. Begin by taking $\mathbf{u}^1 = \mathbf{v}^1$. For $j = 2, ..., N$, let

$$\mathbf{u}^j = \mathbf{v}^j - \frac{\mathbf{u}^1 \cdot \mathbf{v}^j}{\mathbf{u}^1 \cdot \mathbf{u}^1}\mathbf{u}^1 - ... - \frac{\mathbf{u}^{j-1} \cdot \mathbf{v}^j}{\mathbf{u}^{j-1} \cdot \mathbf{u}^{j-1}}\mathbf{u}^{j-1}. \tag{5.1}$$

One obvious problem with this approach is that the calculations become increasingly complicated and lengthy as the j increases. In many of the important examples of orthogonal functions that we study in connection with Sturm-Liouville problems, there is a two-term recursive formula that enables us to generate the next orthogonal function from the two previous ones.