Linearity and linear operators

1 Elements of linear algebra

1.1 Vector spaces and linear combinations

A vector space S is a set of elements - numbers, vectors, functions - that possesses a few important properties. The first is the ability to define the sum of elements in the set and scalar multiplication. For functions and vectors, the way of defining these are obvious. A *linear combination* of elements $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in S$ is any expression of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k$$

where the c's are just scalars. Another property that makes S a vector space is that any linear combination of elements in S is also in S. This is easily verified in most cases - for example, R^n (the set of n-dimensional vectors) and C_0 (the set of continuous functions) are vector spaces.

1.2 Linear transformations and operators

Suppose $\mathbf{A}$ is a $n \times n$ matrix, and $\mathbf{v}$ is a n-dimensional vector. The matrix-vector product $\mathbf{y} = A\mathbf{v}$ can be regarded as a mapping that takes $\mathbf{v}$ as a input and produces the n-dimensional vector $\mathbf{y}$ as an output. More precisely this mapping is a *linear transformation* or *linear operator*, that takes a vector v and "transforms" it into y. The most basic fact about linear transformations and operators is the property of *linearity*. In words, this roughly says that a transformation of a linear combination is the linear combination of the linear transformations. For the matrix A this means that

$$A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2. \tag{1}$$

More generally, if f_1, f_2 are elements of a vector space S, then a linear operator $\mathcal{L}$ is a mapping from S to some other vector space (frequently also S) so that

$$\mathcal{L}(c_1f_1 + c_2f_2) = c_1\mathcal{L}f_1 + c_2\mathcal{L}f_1. \tag{2}$$

1.3 Eigenvalues

Recall that $\mathbf{v} \neq 0, \lambda$ is an eigenvector-eigenvalue pair of $\mathbf{A}$ if

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

Except in rare cases, there are exactly n linearly independent eigenvectors that solve this (the eigenvalues can sometimes be the same for different eigenvectors, however). Why is this a useful fact? Note that if $\mathbf{x}$ is any n-dimensional vector, it can be written in terms of the basis given by the eigenvectors, i.e.

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n. \tag{3}$$

Knowing what the c's are, we can very easily compute the linear transformation A on the vector x (without explicitly carrying out the product) as

$$\begin{aligned} A\mathbf{x} =& A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) \\ =& c_1 A\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 + \ldots + c_n\mathbf{A}\mathbf{v}_n \\ =& c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \ldots + c_n\lambda\mathbf{v}_n. \end{aligned}$$

1.4 Inner products and the adjoint operator

For many vector spaces, one can define an inner product on two elements which produces a scalar output. For $\mathbf{v}_1, \mathbf{v}_2 \in S$, we write the inner product as $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$. We say that $\mathbf{v}_1, \mathbf{v}_2$ are *orthogonal* if $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. The inner product has a nice linearity property: for any vector $\mathbf{v}$ and vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ we have

$$\langle \mathbf{v}, c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \rangle = c_1\langle \mathbf{v}, \mathbf{v}_1 \rangle + c_2\langle \mathbf{v}, \mathbf{v}_2 \rangle + \ldots + c_n\langle \mathbf{v}, \mathbf{v}_n \rangle. \tag{4}$$

In other words, the inner product of a vector and a linear combination is the linear combination of the inner products. The dot product for finite dimensional vectors is the best known example of an inner product, but there are in fact many ways of defining inner products.

Once an inner product is defined, then for any linear transformation or operator $\mathbf{A}$, there is another operator called the *adjoint* of $\mathbf{A}$, written $\mathbf{A}^\dagger$. What defines the adjoint is that for any two vectors $\mathbf{v}_1, \mathbf{v}_2$,

$$\langle \mathbf{A}\mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{A}^\dagger\mathbf{v}_2 \rangle. \tag{5}$$

This definition seems confusing at first because $\mathbf{A}^\dagger$ is not explicitly constructed. You should think of this as "if I find an operator $\mathbf{A}^\dagger$ that satisfies property (5), it must be the adjoint."

If the inner product is just the dot product for n-dimensional (real) vectors, then it turns out that

$$(\mathbf{A}\mathbf{v}_1) \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot (\mathbf{A}^T\mathbf{v}_2),$$

which means that the transpose of $\mathbf{A}$ is same as the adjoint of $\mathbf{A}$. If the inner product were different, this would not be the case.

1.5 Self-adjointness

Sometimes, an operator is its own adjoint, in which case its called *self-adjoint*. Self-adjoint operators have some very nice properties which we will exploit. The most important are

1. The eigenvalues of A are real.

2. The eigenvectors of A are orthogonal to one another.

Suppose A is self-adjoint, and we know its eigenvectors. As in (3), we can try to write any vector $\mathbf{x}$ as a linear combination of the eigenvectors. Actually determining the c's in (3) is usually a hard problem. If A is self adjoint, however, something remarkable happens. Taking the inner product of (3) with any particular eigenvector $\mathbf{v}_k$ and using (4), we have

$$\begin{aligned}\langle \mathbf{x}, \mathbf{v}_k \rangle &= \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n, \mathbf{v}_n \rangle \\ &= c_1\langle \mathbf{v}_1, \mathbf{v}_k \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_k \rangle + \ldots + c_1\langle \mathbf{v}_n, \mathbf{v}_k \rangle \\ &= c_k\langle \mathbf{v}_k, \mathbf{v}_k \rangle\end{aligned} \tag{6}$$

since $\mathbf{v}_k$ is orthogonal to all eigenvectors except itself. Therefore we have a simple formula for any coefficient c_k:

$$c_k = \frac{\langle \mathbf{x}, \mathbf{v}_k \rangle}{\langle \mathbf{v}_k, \mathbf{v}_k \rangle}. \tag{7}$$

In many cases the eigenvectors are rescaled or *normalized* so that $\langle \mathbf{v}_k, \mathbf{v}_k \rangle = 1$, which means that (7) simplifies to $c_k = \langle \mathbf{x}, \mathbf{v}_k \rangle$.

2 Differential linear operators

We can think of differential operators as just linear operators which act on a vector space of functions. This vector space happens to be infinite dimensional (for example $1, x, x^2, \ldots$ are all linearly independent). Fortunately, many ideas about finite dimensional linear operators carry over to infinite dimensions: linearity, eigenvalues, adjoints. For example, the second derivative of a function d^2f/dx^2 can be though of as mapping $f(x)$ to the output $f''(x)$. The basic linearity property (1) is easily verified since

$$\frac{d^2}{dx^2}\Big(c_1 f_1(x) + c_2 f_1(x)\Big) = c_1\frac{d^2 f_1}{dx^2} + c_2\frac{d^2 f_2}{dx^2}.$$

There is a technical point to be made here. We usually have to worry about what set of functions actually constitutes our vector space. For example, we could specify the set of all twice differentiable functions on the real line $C^2[-\infty, \infty]$, or the set of infinitely differentiable functions on the interval $[a, b]$ which equal zero at each endpoint, denoted $C_0^\infty[a, b]$ (the zero subscript typically means the functions are zero on the boundary). Changing the space of functions on the which a differential operator acts can change things like eigenvalues and adjointness properties.

2.1 The superposition principle

All linear differential equations can be written in the form

$$\mathcal{L}f = 0, \quad \text{(homogeneous equations)}, \tag{8}$$

$$\mathcal{L}f = g, \quad \text{(inhomogeneous equations)}. \tag{9}$$

where f is the "solution" – the function to be found. $\mathcal{L}$ is some differential linear operator and g is a given function.

The big advantage of linearity is the *superposition principle*. There are actually two different statements for homogeneous and inhomogeneous equations. For homogeneous equations, if $f_1, f_2, \ldots$ are solutions of $\mathcal{L}f = 0$, then so is any linear combination, since

$$\mathcal{L}(c_1 f_1 + c_2 f_2 + \ldots) = c_1 \mathcal{L} f_1 + c_2 \mathcal{L} f_2 + \ldots = 0 \quad \text{(homogeneous superposition principle)}$$

The permits a divide-and-conquer strategy for solving equations: if we can find a maximal, linear independent set of solutions, we can get all solutions simply by forming linear combinations of these building-blocks.

For inhomogeneous equations, the statement is a little more involved. Suppose that g is itself written as a linear combination of terms $c_1 g_1 + c_2 g_2 + \ldots$, so that the problem to be solved is

$$\mathcal{L}f = c_1 g_1 + c_2 g_2 + \ldots$$

Then if we can find solutions to the (hopefully simpler) inhomogeneous problems $\mathcal{L}f_i = g_i$, then $c_1 f_1 + c_2 f_2 + \ldots$ solves the whole problem:

$$\begin{aligned}\mathcal{L}(c_1 f_1 + c_2 f_2 + \ldots) &= c_1 \mathcal{L} f_1 + c_2 \mathcal{L} f_2 + \ldots \\ &= c_1 g_1 + c_2 g_2 + \ldots = g \quad \text{(inhomogeneous superposition principle)}\end{aligned}$$

2.2 Eigenvalues

Analogous to the matrix case, we can consider an *eigenvalue problem* for a linear operator $\mathcal{L}$, which asks for non-trivial function-number pairs $v(x), \lambda$ which solve

$$\mathcal{L}v(x) = \lambda v(x), \tag{10}$$

It should be noted that sometimes the eigenvalue problem is written instead like $\mathcal{L}v(x) + \lambda v(x) = 0$ (which reverses the sign of λ), but the theory all goes through just the same. The functions which satisfy this are called *eigenfunctions* and each corresponds to an eigenvalue λ. As with eigenvectors, we can rescale eigenfunctions: if $v(x), \lambda$ is an eigenvector-value pair, so is $cv(x), \lambda$ for any number c.

What can we expect to happen in these problems? Since we are working in infinite dimensions, there are typically an infinite number of eigenvalues and eigenfunctions. The set of eigenvalues is also called the *discrete spectrum* (there is another part called the essential spectrum which we will not get into). The eigenfunctions constitute a linearly independent set, so one might think that any function can be written as a (infinite) linear combination of eigenfunctions, analogous to (3). This is not always the case, but when it does happen, the set of eigenfunctions is called *complete*. This property is extremely valuable by allowing functions to be written as linear combinations of simpler functions. The most famous example is the Fourier series, which we discuss later.

2.3 Sturm-Liouville operators

As an illustration, we will consider a class of differential operators called Sturm-Liouville operators that are fairly easy to work with, and come up in the study of partial differential equations. These have the form

$$\mathcal{L} = \frac{d}{dx}\Big(p(x)\frac{d}{dx} + q(x). \tag{11}$$

The notation can be a bit confusing. It means that to apply $\mathcal{L}$ to a function $f(x)$, we put f on the right hand side and distribute terms:

$$\mathcal{L}f = \frac{d}{dx}\Big(p(x)\frac{df}{dx}\Big) + q(x)f(x).$$

The functions p and q are given (in many cases they are just constants). The space $\mathcal{L}$ acts on will be $C_0^\infty[a, b]$, which means that $f(a) = f(b) = 0$ (zero boundary conditions) and f is infinitely differentiable on the interval $[a, b]$ where it lives.

2.4 Inner products and self adjointness

As pointed out earlier, there are many ways of defining an inner product. Given two functions f, g in $C_0^\infty[a,b]$, we define the "L^2" inner product as

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx \tag{12}$$

We can now show that the Sturm-Liouville operator (11) acting on $C_0^\infty[a,b]$ is self-adjoint with respect to this inner product. How does one compute the adjoint of a differential operator? The answer lies in using integration by parts, or in higher dimensions Green's formula (which is just the divergence theorem applied to products). For any two functions f, g in $C_0^\infty[a,b]$ we have

$$\begin{aligned}
\langle \mathcal{L}f, g\rangle &= \int_a^b \frac{d}{dx}\Big(p(x)\frac{df}{dx}\Big)g(x) + q(x)f(x)g(x)dx \\
&= \int_a^b -p(x)\frac{df}{dx}\frac{dg}{dx} + q(x)f(x)g(x)dx \\
&= \int_a^b \frac{d}{dx}\Big(p(x)\frac{dg}{dx}\Big)f(x) + q(x)f(x)g(x)dx \\
&= \langle f, \mathcal{L}g\rangle.
\end{aligned}$$

Integration by parts was used twice to move derivatives off of f and onto g. Thus (comparing to (5))$\mathcal{L}$ is the adjoint of itself. Note that the boundary conditions were essential to make the boundary terms in the integration by parts vanish.

What about a non-self-adjoint operators? The simplest example is $\mathcal{L} = d/dx$ acting on $C_0^\infty[a,b]$. We again compute

$$\langle \mathcal{L}f, g\rangle = \int_a^b \frac{df}{dx}g(x) = -\int_a^b \frac{dg}{dx}f(x)dx = \langle f, -\mathcal{L}g\rangle.$$

It follows that $\mathcal{L}^\dagger = -\mathcal{L}$. Operators like this which are equal the negative adjoints are called *skew symmetric*.

Self-adjoint operators have some properties equivalent to self-adjoint matrices. In particular,

1. The eigenvalues are real.

2. The eigenfunctions are orthogonal to one another.

The second property is extremely useful, especially in connection with Fourier and other series involving orthogonal functions.

2.5 Sturm-Liouville eigenvalue problems

Sturm-Liouville operators (11) on $C_0^\infty[a,b]$ have some other nice properties aside from those of any self-adjoint operator. For the eigenvalue problem

$$\frac{d}{dx}\Big(p(x)\frac{df}{dx}\Big) + q(x)f(x) + \lambda f(x) = 0,$$

(note this has the form $\mathcal{L}f + \lambda f = 0$), the following properties hold:

1. The real eigenvalues can be ordered $\lambda_1 < \lambda_2 < \lambda_3 \ldots$ so that there is a smallest (but not largest) eigenvalue.

2. The eigenfunctions $v_n(x)$ corresponding to each eigenvalue $\lambda_n(x)$ form a complete set, i.e. for any $f \in C_0^\infty[a,b]$, we can write f as a (infinite) linear combination
$$f = \sum_{n=1}^{\infty} c_n v_n(x).$$

Here is the simplest example. Consider the operator $\mathcal{L} = d^2/dx^2$ on the vector space $C_0^\infty[0,\pi]$. The eigenvalue problem reads

$$\frac{d^2 f}{dx^2} + \lambda f(x) = 0. \tag{13}$$

This is just a second order differential equation, and writing down the general solution is easy. Recall that we "guess" solutions of the form $f(x) = \exp(rx)$. Provided $\lambda \geq 0$, we get $r = \pm i\sqrt{\lambda}$, which means that the general solution has the form

$$f(x) = A\cos(\sqrt{\lambda}x) + B\sin(\sqrt{\lambda}x).$$

Not all of these solutions are valid; we require that $f(0) = 0 = f(\pi)$. Therefore $A\cos(0) + B\sin(0) = 0$, so that $A = 0$. The other boundary condition implies $B\sin(\pi\sqrt{\lambda}) = 0$ which is only the case when $\pi\sqrt{\lambda}$ is a multiple of π, or

$$\lambda = n^2, \quad n = 1, 2, 3, \ldots.$$

Corresponding to each of these eigenvalues is the eigenfunction $f_n(x) = \sin(nx), n = 1, 2, 3, \ldots$. Recall that we don't care about the prefactor B since eigenfunctions can always be rescaled.

Properly speaking, we also need to consider the case $\lambda < 0$. We find that the general solution has the form $A\exp(\sqrt{|\lambda|}x) + B\exp(-\sqrt{|\lambda|}x)$ which can't satisfy the boundary conditions unless $A = B = 0$. Be careful, however: there are Sturm-Liouville eigenvalue problems which *do* have non-positive eigenvalues. On the other hand, there is no need to worry about complex eigenvalues because the linear operator is self-adjoint.

2.6 Fourier series

There is a big payoff from fact that the eigenvectors of the previous example are complete: we can write any (!) function in $C_0^\infty[0,\pi]$ as a linear combination of the eigenfunctions of d^2/dx^2. In other words we can write any continuous function with zero boundary values as

$$f(x) = \sum_{n=1}^{\infty} B_n \sin(nx). \tag{14}$$

This is one example of a Fourier series. Other types of Fourier series involve cosines or complex exponentials. These functions are also eigenfunctions of the second derivative operator, but with different boundary conditions on the underlying set of functions on which it acts (remember the technical point earlier: properties of differential operators crucially depend on the vector space in question).

The question that arises is, how do we actually compute the coefficients B_n? We have already answered this question in a more general setting. Because the eigenfunctions $\sin(nx)$ are orthogonal, we can use (6) and (7). For the present case, this is equivalent to taking an inner product of (14) with each eigenfunction $\sin(nx)$. This gives the equivalent of (6), namely

$$B_n = \frac{\langle f(x), \sin(nx)\rangle}{\langle \sin(nx), \sin(nx)\rangle} = \frac{\int_0^\pi f(x)\sin(nx)dx}{\int_0^\pi \sin^2(nx)dx}.$$

It should be emphasized that Fourier coefficient formulas like this one don't need to be memorized. They arise quite simply from the more general idea of finding coefficients of a linear combination of complete, orthogonal set.