Math/Physics 506

Some Ideas of Linear Algebra, Part I: Linear Independence, Span, and Basis

P. A. Perry

A **vector space** is a set of objects ("vectors") which can be added to each other and multiplied by scalars. The operations of addition and scalar multiplication obey the same rules that hold for vectors in two- and three-dimensional space. Examples include

1. The two-dimensional space of vectors of the form $x_1\mathbf{i}+x_2\mathbf{j}$ where x_1 and x_2 are real numbers. We can also write such vectors as pairs (x_1, x_2).

2. The three-dimensional space of vectors of the form $x_1\mathbf{i}+x_2\mathbf{j}+x_3\mathbf{k}$ where x_1, x_2, and x_3 are real numbers. We can also write such vectors as triples (x_1, x_2, x_3).

3. The n-dimensional space of vectors of the form $(x_1, x_2, \cdots, x_n)$ where x_1, x_2, etc., are all real numbers. Mathematicians sometimes call this vector space $\mathbf{R}^n$. If $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$, the vector sum of x and y is the vector $x + y = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)$. If λ is a real number, the scalar multiple λx is the vector $(\lambda x_1, \cdots, \lambda x_n)$.

4. The space of solutions to certain differential equations. Consider, for example, the differential equation $y'' + y = 0$. Two independent solutions of this differential equation are given by $y_1(x) = \cos(x)$ and $y_2(x) = \sin(x)$. Any solution $y(x)$ is a combination $y(x) = c_1y_1(x) + c_2y_2(x)$ of these two special solutions. In the language of vector spaces, we can say that the set of all solutions to this differential equation is a two-dimensional vector space: the "vector" $y(x) = c_1y_1(x) + c_2y_2(x)$ can be represented by the "coordinates" c_1 and c_2.

5. "Spaces" of functions. Consider the set of all (continuous) functions on the interval $[0,1] = \{x : 0 \leq x \leq 1\}$. Think of a function f as a "vector" with

infinitely many components $f(x)$, one for each $x \in [0,1]$. If f and g are any two functions in this space, the new function $h = f + g$ is just the "vector sum" $h(x) = f(x) + g(x)$, and the "scalar multiple" λf is just the function $w(x) = \lambda f(x)$. Thus, the set of continuous functions on $[0,1]$ is also a vector space (any other interval will also work—there is nothing special about $[0,1]$).

Examples 4 and 5 suggest that ideas of linear algebra have application to many other situations besides vectors in two- and three-dimensional space. We will see how true this is when we study ordinary and partial differential equations in Math/Physics 507!

A set of vectors $\mathbf{x}_1, \cdots, \mathbf{x}_m$ is called **linearly independent** if no one of the vectors can be expressed as a linear combination of the others. The basis vectors **i**, **j** and **k** in three-dimensional space are linearly independent; the solutions $y_1(x)$ and $y_2(x)$ in the differential equation of Example 4 are also linearly independent.

A set of vectors $\mathbf{x}_1, \cdots, \mathbf{x}_m$ **spans** a vector space if any vector in that space can be expressed as a linear combination of $\mathbf{x}_1, \cdots, \mathbf{x}_m$. For example the **i** and **j** vectors span two-dimensional space, and the vectors y_1 and y_2 in Example 4 span the vector space of solutions to the differential equation $y'' + y = 0$. On the other hand, the vectors **i**, **j**, and $\mathbf{i} + \mathbf{j}$ span two-dimensional space, but there are "too many" vectors in this set. The reason is that the third vector can be expressed as a sum of the first two, so this set of three vectors is not linearly independent.

A **basis** for a vector space is a set of *linearly independent* vectors that spans the space. Thus the vectors **i**, **j**, and **k** are a basis for three dimensional vector space. It turns out that the vectors $\mathbf{e}_1 = (1, 0, \cdots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \cdots, 0)$, $\cdots$, $\mathbf{e}_n = (0, 0, \cdots, 0, 1)$ are a basis for the vector space $\mathbf{R}^n$ referred to in example 3 above. Finally, to give another "differential equations" example, the space of solutions to the differential equation $y'' - y = 0$ (note the change in sign from Example 4) is spanned by the solutions $y_1(x) = e^x$ and $y_2(x) = e^{-x}$; it is also spanned by the solutions $z_1(x) = \cosh(x)$ and $z_2(x) = \sinh(x)$. The **dimension** of a basis is the number of vectors in it. A basic fact about bases of vector spaces is

Theorem *All bases of a given vector space have the same dimension.*

The **dimension** of a vector space is the number of vectors in a basis. If V is an n-dimensional vector space and $\mathbf{e}_1, \cdots, \mathbf{e}_n$ is a basis for V, any vector $\mathbf{x}$ can be written in exactly one way as

$$\mathbf{x} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n.$$

The numbers $\alpha_1, \cdots, \alpha_n$ are the **coordinates** of $\mathbf{x}$ with respect to the basis $\mathbf{e}_1, \cdots, \mathbf{e}_n$.

Here are some examples relating to bases and coordinates:

1. The usual basis for two-dimensional vector space is the vectors $\mathbf{i}$ and bf j. Another one consists of the vectors $\mathbf{e}_1 = \mathbf{i} + \mathbf{j}$ and the vectors $\mathbf{e}_2 = \mathbf{i} - \mathbf{j}$. The vector $\mathbf{x} = 4\mathbf{i} + 2\mathbf{j}$ is also given by $\mathbf{x} = 3\mathbf{e}_1 + \mathbf{e}_2$. Thus the coordinates of $\mathbf{x}$ with respect to $\mathbf{i}$ and $\mathbf{j}$ are $(4, 2)$ while the coordinates of $\mathbf{x}$ with respect to $\mathbf{e}_1$ and $\mathbf{e}_2$ are $(3, 1)$.

2. The two-dimensional space of solutions to $y'' - y = 0$ has basis $y_1(x) = e^x$ and $y_2(x) = e^{-x}$. The vector y satisfying $y(0) = 1$, $y'(0) = 0$ is expressed uniquely as $y(x) = \frac{1}{2}y_1(x) + \frac{1}{2}y_2(x)$. In the basis $z_1(x) = \cosh(x)$, $z_2(x) = \sinh(x)$, this same vector is expressed as $y(x) = z_1(x)$ (recall $\cosh(x) = \frac{e^x + e^{-x}}{2}$!) Thus the coordinates of y with respect to the first basis are $(1/2, 1/2)$, while the coordinates of y with respect to the second basis are $(1, 0)$.

Math/Physics 506

Some Ideas of Linear Algebra, Part II:

Inner Products (Dot Products) and Orthogonality

An **inner product space** is a vector space with an **inner product** (sometimes called a **dot product** or **scalar product**). Here are some examples:

1. (Real three dimensional vector space) If $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$, then the dot product is

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

2. (real n-dimensional real vector space, $\mathbf{R}^n$) If $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$, then the inner product $\mathbf{x} \cdot \mathbf{y}$ or $(\mathbf{x}, \mathbf{y})$ or $\langle \mathbf{x} \mid \mathbf{y} \rangle$, is

$$(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

3. (*complex* n-dimensional vector space, $\mathbf{C}^n$) Vectors in this space are n-tuples of complex numbers. Let $\mathbf{z} = (z_1, \cdots, z_n)$ and $\mathbf{w} = (w_1, \cdots, w_n)$. The inner product of $\mathbf{z}$ and $\mathbf{w}$, denoted $(\mathbf{z}, \mathbf{w})$ or $\langle \mathbf{z} \mid \mathbf{w} \rangle$, is

$$(\mathbf{z}, \mathbf{w}) = \overline{z_1}\, w_1 + \overline{z_2}\, w_2 + \cdots + \overline{z_n}\, w_n.$$

   Here the bar means complex conjugation.

4. (function spaces) If f and g are functions of one variable defined for $0 \leq t \leq 2\pi$, for example, the "inner product" of f and g is

$$(f, g) = \int_0^{2\pi} \overline{f(t)}\, g(t)\, dt.$$

   The bar over $f(t)$ means that we take its complex conjugate, if it is a complex-valued function like e^{it} (in which case $\overline{e^{it}} = e^{-it}$. This inner product is especially important for the study of Fourier series, as we will see.

Inner products are useful because they allow us to introduce "geometric" notions like length and orthogonality. The **length** of a vector $\mathbf{x}$ in an inner product space is defined by

$$|\mathbf{x}| = \sqrt{(\mathbf{x},\mathbf{x})}$$

where $(\mathbf{x},\mathbf{y})$ is the inner product appropriate to that space. Returning to the above examples:

1. The length of $\mathbf{x} = (x_1, x_2, x_3)$ is given by

$$|\mathbf{x}| = \sqrt{(\mathbf{x},\mathbf{x})} = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

2. The length of $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ is

$$|\mathbf{x}| = \sqrt{(\mathbf{x},\mathbf{x})} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

3. The length of $\mathbf{z} = (z_1, z_2, \cdots, z_n)$ is

$$|\mathbf{z}| = \sqrt{(\mathbf{z},\mathbf{z})} = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2}.$$

   Note that we take the modulus squared of each complex component z_i.

4. The "length" of a function $f(t)$ defined for $0 \leq t \leq 2\pi$ is

$$\|f\| = \sqrt{\int_0^{2\pi} |f(t)|^2\, dt}.$$

   Note in this case we use double bars for length since $|f|$ means the absolute value or modulus of f, which is something different.

Notice that, in any example, a vector can have length zero if and only if it is the zero vector. (A function with "length" zero satisfies $\int |f(t)|^2\, dt = 0$, which can only happen if $f(t) = 0$ for every t).

Two vectors $\mathbf{v}$ and $\mathbf{w}$ are **orthogonal** if $(\mathbf{v},\mathbf{w}) = 0$. In three-dimensional space, this implies that the two vectors are perpendicular. In the other spaces, we *define* orthogonality using the inner product, but still keep the picture in mind of perpendicular vectors in three-dimensional space. Here are some examples to check:

1. The vectors $(-1,0,1)$ and $(0,1,0)$ in three-dimensional space are orthogonal.

2. The vectors $(1,1,\cdots,1)$ (all 1's in every entry) and $(-1,1,0,\cdots,0)$ in real n-dimensional vector space are orthogonal.

3. The vectors $(i,1)$ and $(i,-1)$ in complex two-dimensional space are orthogonal (be careful about taking complex conjugates!)

4. The functions $\cos(t)$ and $\sin(t)$ defined for $0 \le t \le 2\pi$ are orthogonal: inotherwords, their inner product

$$\int_0^{2\pi} \cos(t)\sin(t)\,dt$$

is zero.

Now that we know what length and orthogonality are, we can talk about orthonormal sets of vectors. These are vectors which, like the **i**, **j**, and **k** vectors in real three-dimensional space, have unit length and are mutually perpendicular. More precisely, a set of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ is an **orthonormal set** if $(\mathbf{v}_i, \mathbf{v}_j)$ equals 0 if $i \neq j$, and 1 if $i = j$. Here are some examples for each of our vector spaces.

1. The vectors **i**, **j**, **k** in three-dimensional space.

2. The vectors $\mathbf{e}_1 = (1,0,\cdots,0)$, $\mathbf{e}_2 = (0,1,0,\cdots,0)$, $\cdots$, $\mathbf{e}_i = (0,0,\cdots,1,\cdots,0)$ (the 1 is in the ith slot), $\cdots$, $\mathbf{e}_n = (0,0,\cdots,0,1)$ in real n-dimensional vector space.

3. The same set of vectors in complex n-dimensional vector space.

4. The functions $\phi_n(t) = \frac{1}{\sqrt{2\pi}}e^{int}$ for $0 \le t \le 2\pi$, where $n = 0,1,2,\cdots$. Inotherwords, the "inner product"

$$(\phi_n, \phi_m) = \int_0^{2\pi} \frac{1}{\sqrt{2\pi}}e^{-int}\frac{1}{\sqrt{2\pi}}e^{imt}\,dt$$

is zero if $n \neq m$, and 1 if $n = m$ (note: the reason for the $-$ sign in e^{-int} is the complex conjugation in the inner product). This set of orthonormal "vectors" is very important in the theory of Fourier series.

An **orthonormal basis** for a vector space is a basis of orthonormal vectors. In fact, in *all* of the examples of orthonormal sets just given (including the last one about functions!), the sets are really orthonormal bases for the corresponding vector spaces.

An important technique you should know is the **Gram-Schmidt Orthonormalization Process**. It is a recipe for taking a set of linearly independent vectors (remember part I!), and manufacturing a set of orthonormal vectors spanning the same space. Before we describe it, it is important to note the following recipe for computing the component of a given vector, say $\mathbf{w}$, in the direction of a unit vector $\mathbf{u}$: it is $(\mathbf{u}, \mathbf{w})\mathbf{u}$, since $(\mathbf{u}, \mathbf{w})$ gives the right magnitude and $\mathbf{u}$ gives the right direction. We can check this to be true for real three-dimensional vector space, and we *define* it to be true for the other vector spaces!

Here's how the Gram-Schmidt process works. Suppose I'm given a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ which are all linearly independent; that is, no one vector is a linear combination of any of the others. I manufacture a set of orthonormal vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\cdots$, $\mathbf{u}_n$ from these as follows:

Step 1 Let $\mathbf{u}_1 = \mathbf{v}_1/|\mathbf{v}_1|$. This is easy.

Step 2 Compute $\mathbf{w}_2 = \mathbf{v}_2 - (\mathbf{u}_1, \mathbf{v}_2)\mathbf{u}_1$. This is $\mathbf{v}_2$ with the component parallel to $\mathbf{u}_1$ taken out. The only trouble with $\mathbf{w}_2$ is that it isn't necessarily normalized, so we let $\mathbf{u}_2 = \mathbf{w}_2/|\mathbf{w}_2|$.

Step i Compute

$$\mathbf{w}_i = \mathbf{v}_i - (\mathbf{u}_1, \mathbf{v}_i)\mathbf{u}_1 - \cdots - (\mathbf{u}_{i-1}, \mathbf{v}_i)\mathbf{u}_{i-1}.$$

That is, take out all of the components of $\mathbf{v}_i$ in the $\mathbf{u}_1$ through $\mathbf{u}_{i-1}$ directions. Luckily we have already computed $\mathbf{u}_1$ through $\mathbf{u}_{i-1}$! As in step 2, we normalize to make $\mathbf{u}_i = \mathbf{w}_i/|\mathbf{w}_i|$.

We continue in the same way until we reach **Step n** and we are done.

Here are some exercises involving the Gram-Schmidt method.

1. Find a set of orthonormal vectors that span the plane containing the vectors (1,0,1) and $(1,1,0)$.

2. Consider the polynomials $v_0(t) = 1$, $v_1(t) = t$, and $v_2(t) = t^2$, as "vectors" defined for $-1 \leq t \leq 1$. Define an inner product on these vectors by

$$(v, w) = \int_{-1}^{1} v(t)w(t)\, dt.$$

Find the corresponding orthonormal vectors $u_0(t)$, $u_1(t)$, $u_2(t)$. Up to normalization, these are actually the first three Legendre polynomials, important in the theory of Laplace's equation with cylindrical symmetry.

Math/Physics 506
Some Ideas of Linear Algebra, Part III:
Matrices and Linear Transformations

An **m × n matrix** A is an array of numbers

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The numbers a_{ij} are called the **entries** or **matrix elements** of A. It is customary to denote a matrix by a capital letter, like C, and its entries by the corresponding lower case letter, like c_{ij}. An equation like $C = \{c_{ij}\}$ means that C is the matrix whose entries are the numbers c_{ij}. Two matrices A and B, both $m \times n$, can be added give a new matrix $C = A + B$, where $c_{ij} = a_{ij} + b_{ij}$. If A is an $m \times r$ matrix and B is an $r \times n$ matrix, the $m \times n$ matrix $C = AB$ is defined by the equation

$$c_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}.$$

A **linear transformation** is a map T between vector spaces that obeys the following rules:

(a) For any two vectors $\mathbf{x}$ and $\mathbf{y}$, $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$

(b) For any vector $\mathbf{x}$ and any scalar λ, $T(\lambda\mathbf{x}) = \lambda T(\mathbf{x})$.

We sometimes write $T\mathbf{x}$ for $T(\mathbf{x})$.

Here are some examples of linear transformations:

1. (A linear transformation on three-dimensional vector space) The map

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} 3x_1 - 2x_2 \\ x_1 + 5x_3 \\ x_2 - x_3 \end{pmatrix}$$

   is a linear transformation that takes vectors in $\mathbf{R}^3$ to vectors in $\mathbf{R}^3$. Of course, one can write this transformation using matrix multiplication as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} 3 & -2 & 0 \\ 1 & 0 & 5 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

   or $T(\mathbf{x}) = A\mathbf{x}$ where A is the 3×3 matrix above.

2. (Linear transformations from $\mathbf{R}^n$ to $\mathbf{R}^m$) If A is any $m \times n$ matrix whose entries a_{ij} are all *real* numbers, the formula

$$T(\mathbf{x}) = A\mathbf{x}$$

   defines a linear transformation which maps a vector $\mathbf{x} \in \mathbf{R}^n$ into a vector $A\mathbf{x} \in \mathbf{R}^m$. To be more explicit, if we let $\mathbf{y} = T\mathbf{x}$ and $\mathbf{y} = (y_1, \cdots, y_n)$, then

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

3. (Linear transformations from $\mathbf{C}^n$ into $\mathbf{C}^m$) If A is an $m \times n$ matrix whose entries a_{ij} are *complex* numbers, the formula

$$T(\mathbf{z}) = A\mathbf{z}$$

   defines a linear transformation which maps a vector $\mathbf{z} \in \mathbf{C}^n$ into a vector $A\mathbf{z} \in \mathbf{C}^m$. If $\mathbf{w} = T(\mathbf{z})$, then we have the formula

$$w_i = \sum_{j=1}^{n} a_{ij} z_j.$$

4. (A linear transformation on function space) Let's consider, for the sake of concreteness, continuous functions $f(t)$ defined for $0 \leq t \leq 1$. Let $K(t,s)$ be a function defined for $0 \leq s,t \leq 1$. Observe that if $f(t)$ is a "vector" with one component $f(t)$ for each value of t, so $K(t,s)$ is a "matrix" with one "entry" for each value of t, s. The formula

$$(Tf)(t) = \int_0^1 K(t,s)f(s)ds$$

defines a linear transformation on functions! Let's check the definition above. If f and g are any two functions, and λ is any scalar, we have

$$\begin{aligned} T(f+g)(t) &= \int_0^1 K(t,s)\,(f(s)+g(s))\,ds \\ &= \int_0^1 K(t,s)f(s)ds + \int_0^1 K(t,s)g(s)ds \\ &= Tf(t) + Tg(t) \end{aligned}$$

and

$$\begin{aligned} T(\lambda f)(t) &= \int_0^1 K(t,s)\,\lambda f(s)\,ds \\ &= \lambda \int_0^1 K(t,s)f(s)\,ds \\ &= \lambda T f(t) \end{aligned}$$

Linear operators like this are called **integral operators** and play an important role in the Sturm-Liouville theory of ordinary differential equations which we will study next term.

These example deserve some comment. First of all, notice the similarity between the formula in Example 2,

$$y_i = \sum_{j=1}^{n} a_{ij}x_j$$

and the formula in Example 4,

$$Tf(t) = \int_0^1 K(t,s)\,f(s)\,ds.$$

In the second, the summation over $j = 1, \cdots, n$ (summation over components of $\mathbf{x}$) is replaced by integration over s from 0 to 1 (integration over "components" $f(s)$ of f), and the matrix a_{ij} is replaced by the function $K(t,s)$, sometimes called the **integral kernel** of the operator T. In a very real sense, integral operators are to functions what matrices are to vectors. We will return to this point later.

Secondly, the above examples suggest that there is a close connection between linear transformations and matrices (or, in the case of Example 4, integral kernels). In fact, *every* linear transformation on vectors can be represented by a matrix, and there is a simple way of finding it. First, let's consider the case of linear transformations T acting from $\mathbf{R}^n$ to $\mathbf{R}^m$. How can we find the matrix A that represents T? That is, how can we find the matrix elements a_{ij}?

Before we give the answer, we need a little bit of notation. For real n-dimensional vector space $\mathbf{R}^n$, let $\mathbf{e}_1, \cdots, \mathbf{e}_n$ be basis vectors where $\mathbf{e}_i$ is a vector with ith component 1 and all other components zero. This basis is sometimes called the **standard basis** or the **usual basis** for the vector space $\mathbf{R}^n$. In $\mathbf{R}^3$, for example, the basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ is just the vectors **i**, **j**, **k** by a different name.

Suppose we now have a linear transformation T that takes vectors from $\mathbf{R}^n$ into $\mathbf{R}^m$. Any vector $\mathbf{x} \in \mathbf{R}^n$ can be expressed in the form

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n, \tag{1}$$

and by the linearity of T,

$$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + \cdots + x_n T(\mathbf{e}_n).$$

Thus, we can compute $T(\mathbf{x})$ for *any* $\mathbf{x}$ if we just compute $T(\mathbf{e}_j)$ for all of the vectors $\mathbf{e}_j$ in the standard basis for $\mathbf{R}^n$. Each $T(\mathbf{e}_j)$ is a vector in $\mathbf{R}^m$, so we can write, for some numbers a_{ij}, $i = 1, \cdots, m$,

$$T(\mathbf{e}_j) = a_{1j}\mathbf{e}_1 + \cdots + a_{mj}\mathbf{e}_m. \tag{2}$$

(remember $\mathbf{e}_1, \cdots, \mathbf{e}_m$ is the standard basis for $\mathbf{R}^m$). Then, putting together all of these calculations, we get

$$T(\mathbf{x}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \mathbf{e}_i. \tag{3}$$

That is, *the matrix* $A = \{a_{ij}\}$ *is the matrix of the linear transformation* T.

There is a more 'user-friendly' formula for the matrix elements a_{ij}, and it relates to inner products. Observe that the component of $T\mathbf{x}$ in the $\mathbf{e}_i$ direction is

$$\langle \mathbf{e}_i \mid T\mathbf{x} \rangle;$$

setting $\mathbf{x} = \mathbf{e}_j$ and using equation (3) to compute $T\mathbf{x}$, we find that

$$a_{ij} = \langle \mathbf{e}_i \mid T\mathbf{e}_j \rangle.$$

Thus *the ij-th entry of the matrix A is just the inner product of* $\mathbf{e}_i$ *with* $T\mathbf{e}_j$. One sometimes sees the notation

$$\langle \mathbf{e}_i \mid T \mid \mathbf{e}_j \rangle$$

for this inner product.

As we will see later, it is sometimes useful to use a basis other than the standard one to analyze a linear transformation. In the meantime, it is important to realize that *the matrix of a linear transformation is always computed with reference to a specific choice of basis for the domain and the range.* To illustrate this idea, suppose that T is a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$, and suppose we choose basis vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ for $\mathbf{R}^n$ and $\mathbf{w}_1, \cdots, \mathbf{w}_m$ for $\mathbf{R}^m$. Then the matrix elements of T with respect to these two bases are

$$a'_{ij} = \langle \mathbf{w}_i \mid T\mathbf{v}_j \rangle.$$

Thus if

$$\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n,$$

then $\mathbf{y} = T\mathbf{x}$ is given by

$$\mathbf{y} = y_1\mathbf{w}_1 + \cdots + y_m\mathbf{w}_m$$

product on functions as symmetric and Hermitian matrices have to the inner product on vectors? Recall that the inner product on functions is

$$\langle f \mid g\rangle = \int_0^1 f(t)g(t)\,dt$$

if the functions are real-valued, and

$$\langle f \mid g\rangle = \int_0^1 \overline{f(t)}g(t)\,dt$$

if the functions are complex-valued. If T is a symmetric integral operator, then

$$\begin{aligned}
\langle Tf \mid g\rangle &= \\
&= \int_0^1 \left(\int_0^1 K(t,s)f(s)\,ds\right) g(t)\,dt \\
&= \int_0^1 f(t)\left(\int_0^1 K(t,s)g(t)\,dt\right) ds \\
&= \int_0^1 f(t)\left(\int_0^1 K(s,t)g(t)\,dt\right) ds \\
&= \langle f \mid Tg\rangle,
\end{aligned}$$

so a symmetric integral operator really does behave just like a symmetric matrix. You should compare this calculation with the one that appears on page 7 for symmetric matrices; note that the steps correspond exactly!

You should prove to yourself that if T is a Hermitian integral operator and f and g are complex-valued functions, the same result holds.

Math/Physics 506
Some Ideas of Linear Algebra, Part IV:
Eigenvalues and Eigenvectors

Some Facts about Eigenvectors

Recall that an **eigenvector** of an $n \times n$ matrix A is a vector $\mathbf{x}$ that satisfies the equation

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

for some scalar λ, which is called an **eigenvalue** of the matrix A. In general, λ may be a complex number, even if the matrix elements of A are all real. Geometrically, an eigenvector gives a direction in space in which the linear transformation associated to the matrix A "stretches" vectors by the factor λ. In a moment, we will show how to use determinants to find eigenvalues and eigenvectors. First consider the following example: let A be the matrix

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let T be the linear transformation in three-dimensional vector space with matrix A with respect to the $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ basis. It is not difficult to see that the vectors

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{i} + \mathbf{j} \\ \mathbf{e}_2 &= \mathbf{i} - \mathbf{j} \\ \mathbf{e}_3 &= \mathbf{k} \end{aligned}$$

are eigenvectors of A; in fact,

$$\begin{aligned} T(\mathbf{e}_1) &= 3\mathbf{e}_1 \\ T(\mathbf{e}_2) &= \mathbf{e}_2 \\ T(\mathbf{e}_3) &= \mathbf{e}_3. \end{aligned}$$

Thus, if we use $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ as a basis for three-dimensional vector space, the matrix of T with respect to this new basis, A', is very simple:

$$A' = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Along the $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ axes, T "stretches" vectors by a fixed amount. These axes are sometimes called the **principle axes** of the linear transformation T.

In this case, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ are a set of orthonormal vectors that form a basis for three-dimensional vector space. The reason is that the original matrix A is symmetric. In fact, **symmetric and Hermitian matrices always have a complete set of orthonormal eigenvectors.**

We now want to understand why this is true. To save us some work, notice that any symmetric matrix with real matrix elements is necessarily Hermitian (think about this!), so anything we find out about Hermitian matrices will also be true about symmetric matrices. Henceforth, we'll consider Hermitian matrices only.

First, let's do something easy. Suppose that A is a Hermitian matrix (so its entries may be complex, and its eigenvectors may be complex as well). Let's *suppose* that we are given eigenvectors, $\mathbf{z}$ and $\mathbf{w}$, of A, with eigenvalues λ and μ. What can we find out about these eigenvalues and eigenvectors using the fact that A is Hermitian? We can easily deduce

Fact 1 The eigenvalues must be real. To see this, note that

$$\begin{aligned} \langle \lambda\mathbf{z} \mid \mathbf{z}\rangle &= \langle \mathbf{z} \mid \lambda\mathbf{z}\rangle \\ &= \langle \mathbf{z} \mid A\mathbf{z}\rangle \\ &= \langle A\mathbf{z} \mid \mathbf{z}\rangle \\ &= \langle \lambda\mathbf{z} \mid \mathbf{z}\rangle \\ &= \overline{\lambda}\langle \mathbf{z} \mid \mathbf{z}\rangle \end{aligned}$$

which implies that $\overline{\lambda} = \lambda$. Hence λ is real.

Fact 2 Eigenvectors belonging to different eigenvalues are orthogonal. Again, we simply calculate:

$$\begin{aligned}\langle \mathbf{z} \mid A\mathbf{w} \rangle &= \lambda\langle \mathbf{z} \mid \mathbf{w} \rangle \\ &= \langle A\mathbf{z} \mid \mathbf{w} \rangle \\ &= \mu\langle \mathbf{z} \mid \mathbf{w} \rangle\end{aligned}$$

(remember that the eigenvalues are real, so $\overline{\mu} = \mu$), from which we get the equation

$$(\lambda - \mu)\langle \mathbf{z} \mid \mathbf{w} \rangle = 0$$

which can be true only if either $\lambda = \mu$ or $\langle \mathbf{z} \mid \mathbf{w} \rangle = 0$.

It follows that, if we can find enough eigenvectors to make a basis, we can make them an orthogonal basis, just as in the example above. This is because the ones corresponding to different eigenvalues are "automatically" orthogonal, and we can make the ones corresponding to the same eigenvalue we can orthogonalize using the Gram-Schmidt Process. But how to we find eigenvectors of a given matrix in the first place? And, how do we know that we can find "enough" to make a basis?

How to Find Eigenvectors

Recall that, if A is an $n \times n$ matrix, the equation $A\mathbf{x} = \mathbf{0}$ has a nonzero solution if and only if the determinant of A is zero. (If the determinant of A is nonzero, then A is invertible so that the only possible solution is $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$.) This simple observation will help us find the eigenvalues and eigenvectors of $n \times n$ matrices, whether Hermitian or not.

Rewrite the eigenvalue equation

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

in the form

$$(A - \lambda I)\mathbf{x} = 0.$$

Here I means the $n \times n$ **identity matrix** whose matrix elements are 0 if $i \neq j$ and 1 if $i = j$. This equation has a nonzero vector $\mathbf{x}$ as a solution if, and only if,

$$\det(A - \lambda I) = 0.$$

This equation is called the **characteristic equation** for the matrix A. The function $\det(A-\lambda I)$ is a function of λ called the **characteristic polynomial**. Here are some examples.

1. Let
$$A = \begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix}.$$
The characteristic polynomial of A is $(1-\lambda)(3-\lambda)-6$ and, simplifying, the characteristic equation is $\lambda^2-3\lambda-3=0$. Notice that this equation has complex roots!

2. Let
$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$
The characteristic polynomial is $(2-\lambda)^2-1=0$ and, simplifying, the characteristic equation is $\lambda^2-4\lambda+3=0$. Its roots are $\lambda=1$ and $\lambda=3$.

3. Let
$$A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 1 \end{pmatrix}.$$
Notice that A is a symmetric (hence Hermitian) matrix. The characteristic polynomial for A is $(6-\lambda)\left[(1-\lambda)^2-9\right]$ so the characteristic equation is $(6-\lambda)(\lambda+2)(\lambda-4)=0$.

The eigenvalues are roots of the characteristic equation. Given an eigenvalue, how do we find eigenvectors? We must solve the linear equation $(A-\lambda I)\mathbf{x}=0$. We will take this up in the next set of notes.

Math/Physics 506

Some Ideas of Linear Algebra, Part V:

Diagonalization of Matrices, Special Matrices

First, let's recall some of the notions that have been introduced already, and state some results about diagonalization of matrices.

For *any* $n \times n$ matrix, the **eigenvalues** of A are the roots of the equation (in the variable λ)

$$\det(A - \lambda I) = 0.$$

A vector $\mathbf{x}$ is called an **eigenvector** of A if $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ.

If we can find a linearly independent set of eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ with eigenvalues $\lambda_1, \cdots, \lambda_n$, then we can diagonalize the matrix A as follows. Let S be the $n \times n$ matrix whose columns are the eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$, and let Λ be a diagonal matrix with diagonal entries $\lambda_1, \cdots, \lambda_n$. Then

$$A = S\Lambda S^{-1}$$

and

$$\Lambda = S^{-1}AS.$$

To understand these two equations, recall that S is a change-of-basis matrix. That is, if a given vector $\mathbf{x}$ has coordinates $(x_1, \cdots, x_n)$ with respect to the usual basis of $\mathbf{R}^n$, and $(x'_1, \cdots, x'_n)$ with respect to the basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$ of $\mathbf{R}^n$, then

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = S \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}.$$

Thus A and Λ are the *same* linear transformation represented in two *different* bases.

Notice that, for a general square matrix A, there is no guarantee that we can find a basis of eigenvectors. There are some special kinds of matrices for which this is guaranteed.

1. A **symmetric** matrix is a matrix A such that $A = A^T$, i.e., A is its own transpose. The eigenvalues of a symmetric matrix with real matrix elements are always real, and we can find a complete set of orthonormal eigenvectors. If S is the matrix whose columns are the orthonormal eigenvectors, then S is an **orthogonal matrix**; that is, $S^T S = I$, where I is the identity matrix.

2. A **Hermitian** matrix is a matrix A such that $A = A^\dagger$, i.e., A is its own Hermitian conjugate. The eigenvalues of a Hermitian matrix are always real, and such a matrix has a complete set of eigenvectors which are orthogonal *in the inner product on the complex vector space* $\mathbf{C}^n$. If U is the matrix whose columns are the orthonormal eigenvectors of A, then U is a unitary matrix: that is, $U^\dagger U = I$, where I is the identity matrix.

It is worth pointing out that orthogonal matrices have a special connection to the inner product on $\mathbf{R}^n$, and unitary matrices have a special connection to the inner product on $\mathbf{C}^n$. To discuss this special connection, we first note two simple identities.

1. If $\mathbf{x}$ and $\mathbf{y}$ are any two vectors in $\mathbf{R}^n$, and A is any $n \times n$ matrix with real matrix elements, then

$$\langle A\mathbf{x} \mid \mathbf{y} \rangle = \langle \mathbf{x} \mid A^T\mathbf{y} \rangle. \tag{1}$$

2. If $\mathbf{z}$ and $\mathbf{w}$ are any two vectors in $\mathbf{C}^n$, and A is any $n \times n$ matrix with complex matrix elements, then

$$\langle A\mathbf{z} \mid \mathbf{w} \rangle = \langle \mathbf{z} \mid A^\dagger\mathbf{w} \rangle. \tag{2}$$

What is so special about orthogonal and unitary matrices?

1. A matrix S with real entries is orthogonal if, and only if

$$\langle \mathbf{x} \mid \mathbf{y} \rangle = \langle S\mathbf{x} \mid S\mathbf{y} \rangle \tag{3}$$

for any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^n$. Inotherwords, S preserves lengths of vectors and the angles between them. Therefore, orthogonal matrix is "essentially" a rotation of coordinates in real vector space.

2. A matrix U with complex entries is unitary if, and only if,

$$\langle \mathbf{z} \mid \mathbf{w} \rangle = \langle U\mathbf{z} \mid U\mathbf{w} \rangle \tag{4}$$

for any two vectors $\mathbf{z}$ and $\mathbf{w}$ in $\mathbf{C}^n$. Inotherwords, U preserves "lengths" and "angles" for complex vectors. Therefore, a unitary matrix is "essentially" a rotation of coordinates in complex vector space.

To see why equation (3) is true for a matrix with $S^T S = I$, use equation (1):

$$\begin{aligned} \langle S\mathbf{x} \mid S\mathbf{y} \rangle &= \langle \mathbf{x} \mid S^T S\mathbf{y} \rangle \\ &= \langle \mathbf{x} \mid \mathbf{y} \rangle \end{aligned}$$

since $S^T S = I$.

To see why equation (4) is true for a matrix with $U^\dagger U = I$, use equation (2):

$$\begin{aligned} \langle U\mathbf{z} \mid U\mathbf{w} \rangle &= \langle \mathbf{z} \mid U^\dagger U\mathbf{w} \rangle \\ &= \langle \mathbf{z} \mid \mathbf{w} \rangle \end{aligned}$$

since $U^\dagger U = I$.

Thus, **symmetric matrices can be diagonalized by a rotation of axes in $\mathbf{R}^n$**, and **Hermitian matrices can be diagonalized by a rotation of axes in $\mathbf{C}^n$.**

What about matrices that aren't symmetric or Hermitian? Here the story is a bit more complicated. Eigenvalues may be complex, and eigenvectors need not be orthogonal. Moreover, as the following simple example shows, there might not even be a complete set of eigenvectors.

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then A has characteristic equation $\lambda^2 = 0$ so $\lambda = 0$. However, the only solutions to the eigenvalue equation

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

are all multiples of the vector $(1,0)$. This is a one-dimensional subspace of $\mathbf{R}^2$! In this case we complete the basis by looking for **generalized eigenvectors** which solve the equation $(A-\lambda I)^2\mathbf{x}=0$. Since $A^2=0$, the zero matrix, this equation has plenty of solutions! In particular, the vector space $\mathbf{R}^2$ is spanned by the eigenvector $(0,1)$ and the generalized eigenvector $(1,0)$.

For a general matrix A with real entries, there is a basis of $\mathbf{R}^n$ consisting of eigenvectors and generalized eigenvectors. We won't discuss this general case except in some specific examples.