Linear Algebra for Semantic Representation and Information Retrieval

Dominic Widdows

March 30, 2001

In recent times vectors and linear algebra have become more prominent in linguistics, especially in information retrieval. Because of the comparative novelty of vectors and matrices in linguistic applications, and the widespread teaching of linear algebra in almost every science degree (presenting a sharp contrast with the use and teaching of logic), there appears to be no concise yet thorough introduction to 'what a linguist should know about algebra'.

This is an unfortunate state of affairs, for at least three reasons. The first is practical: the power of modern computers has made the manipulation of huge lists of numbers a viable process, and a vector is the abstract way to think about a list of numbers. In our case these lists are used to record the number of times a particular word appears in different documents, or near a predetermined 'content-bearing word'.

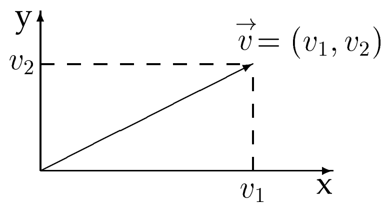
The second reason is historical. The work of Al-Khwarizmi, Descartes, Newton, Einstein and others has enabled a process whereby scientific problems are modeled and manipulated using algebraic structures and numerical equations. An algebraic description of all sorts of spaces and frameworks forms the language in which modern science has been written and technology developed.

Thirdly, vector spaces are the simplest framework we have for developing continuous models. We can generalize many discrete formalisms by allowing for interpolation: for example, by considering values in the interval $[0,1]$ rather than just the binary set $\{0,1\}$. As is common is mathematics, just by dropping a particular constraint we can explore a much richer abstraction.

The purpose of this article is to describe some of the most fundamental algebraic ideas which are used in information retrieval and more recently in latent semantic analysis. I hope to remove some of the mystery behind statements such as "Suppose $\vec{d} = (d_1, \ldots, d_n) \in \mathbb{R}^n$ is the vector of document d", "We project a vector v to a subspace U of dimension k", to explain why a matrix is a mapping between vector spaces and also a list of vectors all at the same time, what a 'metric' is, and how we can talk about the 'angle' between 2 lists of 1000 numbers.

1 Vectors

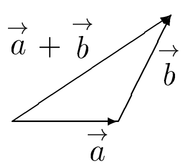
The formal description of a vector was first used to encapsulate our intuitive notion of how points in 2 and 3 dimensional space are structured and how they relate to one another. It is generally believed that the word 'vector' was first used in its modern meaning by Sir William Rowan Hamilton in his work on quaternions, the imaginary quaternion $a\mathbf{i} + b\,\mathbf{j} + c\mathbf{k}$ corresponding to the vector in $\mathbb{R}^3$ (3-dimensional space) with coordinates (a, b, c). In school you will almost certainly have used vectors without necessarily knowing it, when talking about the x- and y-coordinates of a point in the plane. The basic picture is as follows:



Already we see some fundamental notions developing. Our space is first and foremost a set of points : the points are the 'elements' of this set. [1]

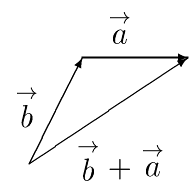
Once you've chosen suitable coordinate axes (this is called "choosing a basis"), vectors in the space can be described by a list of numbers called coordinates. It is very important to distinguish these two ideas — you can have very different coordinate representations for the same vector, just by rotating or rescaling the coordinate axes whilst leaving the vectors strictly alone.

So, what is fundamental about the structure of these points under which they form a 'vector space'? To answer this question, we will look at some examples in a 2 dimensional plane. The only special thing about 2 dimensions here is that it is the most expressive way to draw things on a piece of paper. The more you can think of points as just being elements of an abstract set satisfying certain properties, the better.



Think of a vector as a journey in some direction, and call this journey vector $\vec{a}$. If we have another journey-vector $\vec{b}$, one option is to go along journey $\vec{a}$ and then, starting from where $\vec{a}$ finishes, undertake journey $\vec{b}$. The vector representing the combined journey is naturally called $\vec{a} + \vec{b}$, the *sum* of $\vec{a}$ and $\vec{b}$.

Instead of this we could have chosen to go along journey-vector $\vec{b}$ and then along vector $\vec{a}$, giving vector $\vec{b} + \vec{a}$. One fundamental property of the plane is that these two journeys have the same destination: $\vec{a} + \vec{b}$ and $\vec{b} + \vec{a}$ are two ways of writing the same vector. Thus vector addition is *commutative*.



The process of putting two vectors together like this is of course called vector addition. You should convince yourself that this is a wise generalization of the real number addition operation you learnt as a child. In this context, real numbers behave as '1-dimensional vectors'. Draw yourself a mental picture of how our $a + b$ and $b + a$ journeys appear along a single line, and I hope you'll see what I mean. Another way of describing vector addition is called the 'parallelogram rule'. If you can see why, you're well on the way to understanding what's

[1] Whenever mathematics talks about a 'something-space', the space bit means a set whose elements are called points, and the 'something' describes some structure which the space has, for example vector space, topological space and metric space. We will call our points vectors, though please don't generalize and call points in a topological space topologics and certainly don't refer to points in a metric space as metrics!

I recently had a conversation with George Lakoff in which he sternly denounced the validity of this notion of space as a set of points. We will continue as always by appealing to the model's usefulness.

going on.

One thing of which you should convince yourself is that the commutative property does not hold in other spaces in which you could imagine making a journey. Imagine your space is a sphere like the earth and you start on the equator. If you go 1000 miles north and then 1000 miles east, you will end up further east that if you go 1000 miles east, *then* 1000 miles north. This sentence may sound confusing, but if you give it a little thought the next time you go for a walk, you'll hopefully see what I mean. (Here's a clue: the parallel 1000 miles north is a smaller circle than the parallel of the equator.) The amount of variation which can be introduced by adding journeys in different orders is one measure of 'curvature'. Another way of describing this is that it's the amount by which parallel lines can converge or diverge. This way of thinking about curvature is very useful, because it makes sense however many coordinates you have, so you can extrapolate from the observation that the 2-dimensional surface of the earth is curved to realizing that 4-dimensional spacetime might be similarly curved.

Vector spaces are special precisely because they *aren't* curved. They are spaces which satisfy Euclid's famously dodgy fifth axiom, that parallel lines never get any closer or further away from one another however far you go. [2]

That it took mathematicians over 2000 years to realize that not all spaces behave like this is amazing.

1.1 Formal Rules

Let's try and express the rules which most clearly capture our intuitive notion of how vectors behave.

Firstly, we should be able to add them together. We should be able to go along one, then another, and this whole journey should also be represented by a vector. And it should not matter what order we do this in. Formally, we say that vectors form a *commutative group* under addition. This also entails that we must have a 'zero vector': a null vector which doesn't go anywhere or make any contribution under addition.

But there's more to it than this. Any journey we make in a flat plane can be extended or reduced — we have an intuitive notion of 'half as far' or 'three times the distance', which involves the idea that we are going in the same direction. This idea is also contained in Euclid's first and second postulates: that we can interpolate between two points and extend any straight line indefinitely. The modern name for this process is *scalar multiplication*. Any vector can be multiplied by a number which scales its length by a given factor. This results in vector with the same direction but of different *magnitude*. This scalar multiplication must also satisfy special properties. In particular, it doesn't matter whether you do it before or after addition, so for example $3(\vec{a} + \vec{b})$ must always be the same as $3\ \vec{a} + 3\ \vec{b}$. And there must also be a

[2] Euclid's Elements will probably always remain the best place to learn about formal structures. Whilst reinventing the wheel has made many modern theorists famously proud of themselves, it's not big and it's not clever. If you want to truly understand these things, read at least a bit of Euclid. There is a fantastic interactive version on the web at `http://aleph0.clarku.edu/ djoyce/java/elements/toc.html`. Ten minutes of ancient wisdom will do you more good than reading the rest of these notes.

multiplicative identity: a scale factor (normally called 1) which doesn't alter lengths at all.

This just about covers the axioms for a vector space. Here they are in one place.

Definition 1.1 A (real) Vector Space is a set V equipped with two mappings, called addition ($+ : V \times V \to V$) and scalar multiplication ($\times : \mathbb{R} \times V \to V$).

Addition must obey the following axioms:

- (i) Addition is associative, so that for all $a, b, c \in V$, $(a + b) + c = a + (b + c)$.
- (ii) Addition is commutative, so that for all $a, b \in V$, $a + b = b + a$.
- (iii) There is an additive identity element $0 \in V$ such that for all $a \in V$, $v + 0 = 0 + v$.

Scalar Multiplication must obey the following axioms:

- (i) Scalar multiplication is associative, so that for all $\lambda, \mu \in \mathbb{R}$ and for all $a \in V$, $(\lambda\mu)a = \lambda(\mu a)$.
- (ii) Scalar multiplication is commutative over addition in $\mathbb{R}$, so that for all $\lambda, \mu \in \mathbb{R}$ and for all $a \in V$, $(\lambda + \mu)a = \lambda a + \mu a$.
- (iii) Scalar multiplication is distributive over addition in V, so that for all $\lambda \in \mathbb{R}$ and for all $a, b \in V$, $\lambda(a + b) = \lambda a + \lambda b$.

All other vector space axioms you will ever see are equivalent to these. For example, you might sometimes see the property of 'closure' referred to, i.e. the constraint that for all $a, b \in V$, $a + b$ is also in V. But we have already stated that by saying that our addition operator maps $V \times V$ to V.

Using multiplication (i) we get that $1(\mu a) = (1\mu)a = \mu a$ for all $\mu \in \mathbb{R}, a \in V$. It follows that scalar multiplication by 1 is an identity operator on V. You can use multiplication (iii) in a similar way to show that scalar multiplication by zero kills all vectors, i.e. for $0 \in \mathbb{R}$ and for all $a \in V$, $0a = 0 \in V$. Try it at home if you like.

There is much here which will take getting used to: I haven't spared you the nitty-gritty. Becoming comfortable with the language is as important and often more difficult than understanding the ideas. [3] As with other languages, I've quite quickly relaxed the formalism and relied on context to tell you what things are and how they are related. I've given up on those fiddly arrows and left Greek versus Roman script to tell you what's a scalar (real number) and what's a vector. This should always be enough to tell you whether addition (written '+') and multiplication (written by juxtaposition) refers to real number or vector operations.

The reasons for this laxity are twofold. First, in all honesty, mathematicians are pretty lazy. Logicians are very energetic and bother in great detail about what something means and precisely what kind of object it is. Mathematicians are often happy to leave details to look after themselves provided the bigger picture is saying something useful. The second is a positive reason — it is generality. Once you relax your

[3] I have heard mathematics described as "a language which takes years to learn with the added drawback that there are no native speakers".

stringent formality you begin to realize how many rich structures can be described by this framework.

Lots of things can be described as vectors that you never suspected of so being. The real numbers themselves are vectors — go ahead and check the axioms if you like. The set $\mathbb{R}^{\mathbb{R}}$ of real-valued functions also forms a vector space. If we have two functions $f, g : \mathbb{R} \to \mathbb{R}$, then $f + g$ is also a function defined by $(f + g)(x) = f(x) + g(x)$, and so is λf defined by $(\lambda f)(x) = \lambda(f(x))$ for all $\lambda, x \in \mathbb{R}$. If however I started emphasizing this point by writing $\vec{f}$ and $\vec{g}$ instead of f and g, you would probably find it confusing rather than helpful and wonder what sort of arcane incantation was being invoked. No, vectors are perfectly normal. Any objects which you can add together and multiply by real numbers in a well-behaved fashion will probably form a vector space. It will come as no surprise that this leads to a variety powerful and useful ideas.

2 Word-Vectors

Rather than go into more theoretical details immediately, we will go straight on to explain the way in which vectors can be used to describe words and their meanings.

The idea is to give each word a set of coordinates which measures how the extent to which the word is related to a particular context. Here is an excerpt from the sort of table we might create:

	base	bass	buy	goal	player	tickets	tunes
Business	14	0	23	9	6	6	2
Music	3	13	8	2	12	8	11
Sport	12	0	7	14	13	10	4

Table 1: A matrix of possible words and contexts

The "contexts" (Business, Music and Sport) are listed on the left, the terms along the top, and the numbers denote the relevance of these terms in the given contexts. Such data can be built up by electronically reading documents about the topics in question. Consider, for example, the following passage (from the Chicago Sunday Times):

> The missing link between the western swing style of Bob Wills and Texas Playboys and the hot jazz of Django Reinhardt and Stephane Grappelli can be found right here in the **music** of Hot Club of Cowtown. The three-piece combo – Whit Smith (guitar), Elana Fremerman (violin) and Matt Weiner (**bass**) – from Austin, Texas, puts its own stamp on old fiddle **tunes** and Tin Pan Ally standards. Johnny Gimble opens. **Tickets**, $16-$20, at the Old Town School (773-728-6000).

A table like the one above can be built up gradually, by adding 1 to each word's 'music' score every time it appears in a document close to the word 'music'. In this way the table records the fact that the words ‘bass’, ‘tunes’ and ‘tickets’ (and the words ‘jazz’, ‘combo’, ’swing’ and many others) often occur in a musical context.

Of course, there are many questions that arise when considering this basic approach, and many decisions that must go into building the model. What should count as a ‘topic’ or ‘context’, and how should a piece of text be assigned to a particular subject in practice? Once assigned, should all occurrences of a term be considered as having equal importance? How does the model deal with words that occur in a variety of different contexts? Some of the most important words such as articles, pronouns and conjunctions can occur in almost any situation, and their content is entirely determined by the words around them. In fact, word vectors sometimes capture this surprisingly well. For example, color words can appear in all sorts of contexts, but often with similar distributions to other color words. As a result their word-vectors are recognizably similar to one another even though their uses are very diverse.

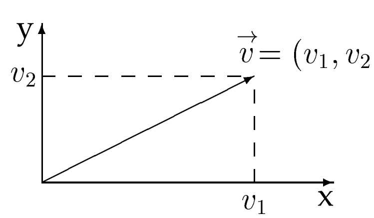
I hope the example above gives some idea of the way in which certain features can be used to assign characteristics to words and their uses. In Table (1) above these features are ‘context-words’, but they could equally well be paragraphs, abstracts, or whole documents. For example, most Information Retrieval systems work by giving each word a ‘document-profile’ according to which documents that word appears in: which inversely gives each document a ‘word-profile’, a characteristic fingerprint of which words appear in the document.

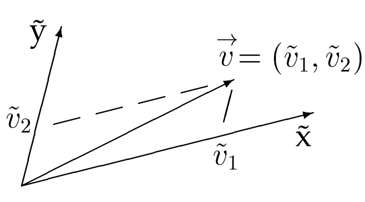
Now, here’s the point. Once words are mapped to lists of numbers, those lists can be added together component by component, and multiplied by any other real number. It is a simple matter to check that these lists, and the operations of list-addition and scalar multiplication, satisfy the axioms of Definition 1.1. In other words, the list of numbers attached to each word can be thought of as vectors, which naturally enough we call *word-vectors*.

2.1 Bases, Coordinates and Dimensions

In order to make sense of these word-vectors, we need to understand a few more basic concepts.

Recall our original picture of a vector v with its coordinates (v_1, v_2). These coordinates are clearly affected by the way we chose our x-axis and y-axis. For example, we could rotate these axes through some angle, and then the coordinates would be different. Nor is there any *a priori* reason why the axes have to be at right angles. Once we take such possibilities into account, we allow a whole variety of different representations.





For example, the picture on the left has a completely different pair of axes. Nonetheless, the coordinates $(\tilde{v}_1, \tilde{v}_2)$ still give a unique, unambiguous description of the vector $\vec{v}$, provided we know which vectors x̃ and ỹ are. So long as these two axes are not parallel, we can get to any point in the plane by going some distance $\tilde{v}_1$ along our x̃-axis and some distance $\tilde{v}_2$ along our ỹ-axis.

This method of using pairs of numbers to refer to points in the plane was invented by Descartes. To this day, the coordinates are called Cartesian coordinates and the technique is called Cartesian geometry.

Important definitions follow.

Definition 2.1 A set $\{u_1, \ldots, u_n\}$ of coordinate axes for a vector space U is called a *basis* for U. Any set of basis vectors must be comprehensive enough so that every vector $v \in U$ can be represented as a *linear combination* of $\{u_1, \ldots, u_n\}$, i.e. there exist scalars $\lambda_i \in \mathbb{R}$ such that

$$v = \lambda_1 u_1 + \lambda_2 u_2 + \ldots \lambda_n u_n.$$

The scalars $(\lambda_1, \ldots, \lambda_n)$ are the called *coordinates* of v with respect to the basis $\{u_1, \ldots, u_n\}$. Since all vectors in U can be built out of the basis $\{u_1, \ldots, u_n\}$, the basis is said to *generate* the space U.

Any set of basis vectors must also be small enough to ensure that this way of writing v is unique — there must be no redundant basis vectors.

One of the key theorems in linear algebra proves that these two conditions — that a basis must be both large enough to give coordinates to all vectors in the space, and small enough so that these coordinates are uniquely determined — are enough to guarantee that every possible basis for the space U has exactly the same number of elements. This magic number n is called the *dimension* of U.

There's a great deal there and I'm not at all surprised if it's left you reeling. I'll try to elucidate matters with a few examples.

Two dimensions — the plane

Perhaps the most familiar example of a vector space (after the real numbers themselves) is the plane. All the drawings we've given so far have been done in a plane, for obvious reasons.

To represent points on the plane, we choose a basis with 2 coordinate axes (as we've done the in some of the examples). Call these vectors $\vec{a}$ and $\vec{b}$. The only condition $\vec{a}$ and $\vec{b}$ have to satisfy is that they must not point in exactly the same direction. (If they did, we'd only be able to represent points on the line that $\vec{a}$ and $\vec{b}$ share: there would be no way of moving laterally.) For example, the vectors $\vec{a}$ and $\vec{b}$ in the diagrams on page 2 form a basis for the plane.

Any point in the plane can be reached by the instructions "go λ times along the vector $\vec{a}$, then go μ times along the vector $\vec{b}$". (The direction may have to be 'backwards', of course, which is described by letting λ or μ be negative.) Thus any vector $\vec{v}$ in the plane can be written as

$$\vec{v} = \lambda \vec{a} + \mu \vec{b},$$

in which case we say that (λ, μ) are the coordinates of the vector $\vec{v}$ with respect to the basis $\{\vec{a}, \vec{b}\}$. It is also the case that λ and μ are completely determined, i.e. they are unique. (You should convince yourself of this — there are several ways to do this.)

Suppose we tried to add a third basis vector $\vec{c}$ somewhere in the plain. All vectors can already be written in terms of just $\vec{a}$ and $\vec{b}$, including $\vec{c}$. It follows that there are a whole variety of ways to write $\vec{v}$ in terms of the set $\{\vec{a}, \vec{b}, \vec{c}\}$. (As well as "go 2 miles north and 1 mile east", you could say "go 1 mile north and $\sqrt{2}$ miles north-east", "go $2\sqrt{2}$ miles north-east -1 mile east", and an infinity of other possibilities.) The representation is no longer unique, so the set $\{\vec{a}, \vec{b}, \vec{c}\}$ is *not* a basis for the plane. It becomes evident that *every* basis for the plane must have precisely 2 independent members.

In order to warrant inclusion, a third basis vector would have to describe some new part of the space that couldn't have been reached by $\vec{a}$ and $\vec{b}$ alone — a third dimension.

Higher dimensions — the vector space $\mathbb{R}^n$

Spaces like this are so important that we give them a special name. The set of all real numbers is written as the letter $\mathbb{R}$. By choosing a suitable basis, every point in the plane can be represented by a pair (λ, μ) of real numbers, which is the same thing as an element of the Cartesian product $\mathbb{R} \times \mathbb{R}$. We call this space $\mathbb{R}^2$ (normally pronounced 'R two' *a la* Star Wars rather than '$\mathbb{R}$ squared' which you might have expected.)

By analogy, three dimensional space can be represented as triples of real numbers in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, and this space is given the name $\mathbb{R}^3$ (pronounced 'R three'). And so on — quadruples of real numbers live in $\mathbb{R}^4$ ('R four') and n-tuples of real numbers live in $\mathbb{R}^n$ ('R to the n').

To make our notation truly general and scalable (i.e. to avoid running out of letters), coordinates and basis vectors are usually written using subscript (or superscript) indices rather than different letters. Instead of calling a basis $\{\vec{a}, \vec{b}\}$, we will usually write $\{e_1, e_2\}$, and instead of writing $\vec{v} = \lambda \ \vec{a} \ + \mu \ \vec{b} \in \mathbb{R}^2$ we will usually use $\vec{v} = \lambda_1 e_1 + \lambda_2 e_2 in \mathbb{R}^2$. This really pays off in higher dimensions where it permits the generalized notation

$$\vec{v} = \lambda_1 e_1 + \ldots + \lambda_n e_n \in \mathbb{R}^n, \tag{1}$$

or just $\vec{v} = \sum_{i=1}^n \lambda_i e_i \in \mathbb{R}^n$. Yet another way of doing this is to write $\vec{v} = (\lambda_1, \ldots, \lambda_n)$, giving only the coordinates assuming that the basis is already agreed upon. (This is probably the way you were taught to represent points in the plane at school, as a pair (x, y), and you were probably never told that this representation depends on choosing the x- and y-axes as a basis.) The coordinate λ_i is sometimes referred to as the "component of $\vec{v}$ in the e_i-direction", or just the i^{th}-*component* or i^{th}-*coordinate* of the vector $\vec{v}$.

Using this notation, let us formally define addition and scalar multiplication in $\mathbb{R}^n$. (We should do this anyway and it will help to get you to get used to the notation.)

Let $\{e_1, \ldots, e_n\}$ be a basis for $\mathbb{R}^n$. Let $\vec{v}_1 = \lambda_1 e_1 + \ldots + \lambda_n e_n$ and $\vec{v}_2 = \mu_1 e_1 + \ldots + \mu_n e_n \in \mathbb{R}^n$ be vectors in $\mathbb{R}^n$. Their sum is defined to be

$$\vec{v}_1 + \vec{v}_2 = (\lambda_1 + \mu_1) e_1 + \ldots + (\lambda_n + \mu_n) e_n.$$

Let $\alpha \in \mathbb{R}$. Scalar multiplication is defined by the equation

$$\alpha \vec{v} = \alpha\lambda_1 e_1 + \ldots + \alpha\lambda_n e_n.$$

A shorthand-way to say this is that vector addition and multiplication in $\mathbb{R}^n$ is carried out 'component by component'.

A simple but very good exercise at this point would be to check that these operations satisfy the vector space axioms of Definition 1.1. (More challenging: doing this relies on fundamental properties of real numbers which you probably take for granted. Which properties of real numbers do you assume, and where do you need to invoke them?)

Choosing Bases

By now you should be aware of several key points. You should have an idea of what vectors are, what bases are, and of coordinates with respect to a given basis.

One critical point to remember is that a vector is not the same as its coordinates: a vector is *represented* by its coordinates. This is easy to overlook, especially since mathematicians often write things like "let $\vec{v} = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$", which assumes that a basis has already been chosen and omits to mention this for the sake of brevity.

In spite of its supposed exactness, mathematics is full of this sort of 'agreed shorthand'. Diligent authors make such abbreviations transparent by writing sentences such as "by abuse of notation we write...". It is important to be on your guard. Often such abuses and fiddles go unnoticed for centuries. Ground breaking research has often been inspired when someone suddenly realizes that things which have been lumped together and assumed to be the same are in fact quite different.

This has certainly been the case in recognizing the difference between a point in space and its coordinate representations. The controversy which so entangled Galileo — whether the earth is at the center of the universe — rested upon just such a misconception. Whether the earth is 'at the center' depends upon choosing a coordinate system for the universe — until you've chosen a basis, the question doesn't really mean anything. But choosing a basis is part of the conceptual model you're using, not part of space itself. The question becomes "is it permissible to choose coordinates which do not treat the earth as the static center of the universe".

This might sound obvious, but it really isn't (not to me at least!). That the universe is independent from any one particular set of coordinates is an insight that only really came to fruition in the work of Albert Einstein. Not only does regarding any one point as 'central' depend upon choosing coordinates: the same is true of regarding any one point as 'still'. It makes no absolute sense to say that a particle is 'at rest': all you can do is to say that a particle is at rest with respect to a particular choice of basis. (This insight is from the special theory of relativity: general relativity goes even further by recognizing that, because gravity causes the universe to curve, there is no way to assign flat basis vectors to spacetime. So setting a basis for space and time is not only a subjective choice: it's also only a local approximation.)

It is important to realize which properties of vectors depend upon coordinates with respect to a particular basis, and which are basis independent. For example the famous equation $E = mc^2$ is none other than the result of realizing that an object's kinetic energy $E = mv^2$ depends on choosing a basis in which to measure its velocity v. Since energy is taken to be a constant, independent of choosing a basis, there should be an amount of energy stored even in an object which appears to be static. Lo and behold, this energy is precisely the amount of energy the object would have if it traveled at the speed of light, which conveniently tells you exactly how much energy is released in a nuclear reaction, when mass is converted into electromagnetic radiation traveling at light speed.

An important (and quite difficult) exercise is to check that even though vector addition and multiplication in $\mathbb{R}^n$ are defined in terms of a particular basis, you actually get the same sum and scalar product *whichever basis you choose*.

2.2 Choosing bases for word-vectors

By now I hope that all this work on bases and coordinates has begun to make sense, and that you can already see where it's leading us in describing word-vectors. Understanding vectors allows us to use a range of simple, versatile ideas which have been thoroughly developed over the past 300 years (and really much longer) and which have proven their ability to throw light on a vast variety of topics which we encounter in the real and abstract worlds.

Refer back to our first word-vectors in Table (1). We can now say a lot more about this table in terms of vectors, bases and coordinates. We have already said that each column of numbers in the table can be described as a vector. Each individual cell in the table is a measure of how much of each context goes into the making of the word-vector. For example we have given the word 'player' the word-vector

$$\overrightarrow{\text{player}} = 6(\text{Business}) + 12(\text{Music}) + 13(\text{Sport}).$$

This notation makes the analogy very clear: it is in exactly the same form as Equation (1) which represents a vector in terms of coordinates and basis vectors. Our 'contexts' have taken the role of basis vectors, and the individual cells in the table are playing the role of coordinates. So the 'music-coordinate' of the word-vector of 'player' is 12. We could also say that "the component of the 'player' vector in the 'music' context is 12".

This motivates some of the most varied and interesting questions we shall encounter. What should our basis vectors be? How should we choose them? Once chosen, how much weight should be given to the occurrence of a particular word near a particular context? Ideally, we will eventually have a model for meaning which is independent of choosing a basis, more general than any one particular corpus of literature, more general even than any one particular language. This dream will take a long time to achieve. In order to approach it we must try to build models by choosing bases and learning coordinates for word vectors in the best possible fashion. Many scientific discoveries tell a story of several different experiments, different frameworks which gradually enabled thinkers to abstract away from the particulars of any one system and to apprehend the underlying phenomena which all the different observations share. In the meantime, we would do well to watch ourselves closely and to note all the different and contestable assumptions we make along the way. Mathematical models can be a window on to the secrets of the universe, but we must avoid the tempting arrogance of thinking that man-made models are the secrets themselves.

The identification of ‘contexts’ in natural language with basis vectors in linear algebra is a very interesting link, upon which the entire success of models such as these rests. There are some good reasons for exploring such a link. To start with, it has proved tremendously successful in determining the relevance of a piece of writing to a request for information. More importantly, linguists often talk about “dimensions of meaning”. In mathematics (and so ultimately in science as a whole), a dimension is *precisely* an independent basis vector in a vector space (or a more generalized curvilinear coordinate, but we won’t go into differential geometry just yet). ‘Are there fundamental dimensions of meaning and is this a useful notion for describing a lexicon’ is *exactly the same question* as ‘Can fundamental meanings be modeled as a set of basis vectors for a space of word-meanings?’

This states the case very strongly and gives rise to several very reasonable doubts. The way vectors are generated from a basis is just by scaling the basis vectors and adding together the results. This is really a very blunt tool for composing fundamental concepts to get new word meanings from old. More generally, vectors are defined to be things that you can add together and multiply by (real) numbers. Words are not. These issues have to be addressed to give any real credence to the vector model. The role of scholarship should be to test such conceptual models as carefully and fairly as possible, not to preach the all-encompassing benefits of ones own approach in the hope that this will lead to recognition. Nonetheless, a few reasons can already be given which suggest that the possibilities of a vector model are well worth exploring, and I will take up this opportunity to begin to describe them.

Much has been made of ‘formal approaches’ in linguistics, yet there are a host of rich and varied abstract structures whose aptitude for describing linguistic phenomena has yet to be explored. Along with set theory and logic, linear algebra is one of the best candidates for preliminary studies. Other models, other spaces, operations other than simple addition and scaling, may provide settings much more suitable for expressing the richness of human communication: but if the simple setting is not explored first it is unlikely that we will ever each the more subtle ones. By the same token, so many formal abstractions are based on linear and algebraic models that if linear algebra turns out to be a complete failure for describing language, it seems very unlikely that any of these other abstract formalisms will be even moderately successful. Take a practical example: you may argue that by ‘dimensions of meaning’ nothing whatsoever is meant to do with vectors and algebra. But *every* mathematical and scientific notion of dimensionality derives from the idea of directions which are somehow independent, and this is encoded in the idea of ‘the number of coordinates needed to specify a point’. If this latter idea has no linguistic significance at all, one would have to give a totally different notion of ‘dimension’ for use in linguistics, which would be at variance with science as a whole.

There are many possible arguments for the relevance of linear algebra for modeling words and their meanings. But the most convincing of all will be to get on with the next steps in the theory and to show how it can be used to provide basic models for some of the ways we use language.