Some stuff about Vector Spaces.

Many rules you have found useful in calculus have one of these 2 forms:

"*The* (something) *of the sum, is the sum of the* (something)*s*."

"*The* (something) *of a constant multiple of* (what you're working with) *is that constant multiple of the* (something) *of* (what you're working with)."

Examples: differentiation, integration, limit of a function, limit of a sequence, "sum" of an infinite series. There are other rules that apply to products or quotients, but those rules have different forms depending on what the (something) is - for example, "limit" and "derivative." And when (something) is "integral," there is *no* rule that expresses the integral of the product or the quotient of 2 functions in terms of the integrals of each of the functions. So addition, and multiplication by a constant - a scalar - are important because they keep showing up, and because they behave so predictably.

The abstract theory known as Linear Algebra is the study of what is "always" true when you work with mathematical objects that can be added and multiplied by scalars, and the study of what is "sometimes" true, in various interesting special circumstances. Our study will be brief.

The objects of our study: vector spaces, linear operations, vectors

A (real) vector space consists of a non-empty set V, the elements of it being called vectors, and 2 operations, addition, defined somehow for *every* pair (v_1, v_2) of elements of V, and scalar multiplication, defined somehow for *every* pair (c, v), where c is a real number, and

v is an element of V. Addition of the pair (v_1, v_2) of elements of V *must* result in an element of V, and that result, called their (vector) sum is expressed in symbols as $v_1 + v_2$. Scalar multiplication of the pair (c, v) of a real number and an element of V *must* result in an element of V, and that result is expressed in symbols as cv. Both operations *must* yield a result that is an element of V: $v_1 + v_2$ and cv both must exist and satisfy the conditions of membership in V. The operations of addition and scalar multiplication, known as the linear operations have to satisfy some more conditions in order for the theory to describe the situations we're used to. Before listing these properties of the linear operations, let's look at a number of examples of things we're used to being able to add and multiply by scalars - and have the result *always* be a thing of the same kind we started with. These will be followed by some *non*-examples.

Examples

1. Matrices of the same size.
2. Polynomials. Addition is defined pointwise: $(P+Q)(x) = P(x) + Q(x)$; scalar multiplication is also defined pointwise, by $(cP)(x) = c(P(x))$.
3. Functions that are differentiable on an interval. Addition is defined pointwise: $(f+g)(x) = f(x) + g(x)$; scalar multiplication is also defined pointwise, by $(cf)(x) = c(f(x))$.
4. Functions that are defined on an interval. Addition is defined pointwise: $(f+g)(x) = f(x) + g(x)$; scalar multiplication is also defined pointwise, by $(cf)(x) = c(f(x))$.

There is a sameness in 2, 3, and 4 that we can exploit later with the idea of a subspace.

5. Sequences. Addition is defined termwise: if $s = \{s_n\}$ and $t = \{t_n\}$ are sequences, the sequence $s + t$ is defined by $(s + t)_n = s_n + t_n$, $n = 1, 2, 3, \ldots$ Scalar multiplication is also defined termwise, by $(cs)_n = c(s_n)$.

6. Series. Addition is defined termwise: if $s = \sum_{n=1}^{\infty} s_n$ and $t = \sum_{n=1}^{\infty} t_n$ are series, the series $s + t$ is defined by $(s + t) = \sum_{n=1}^{\infty} (s_n + t_n)$. Scalar multiplication is also defined termwise, by $cs = \sum_{n=1}^{\infty} c(s_n)$.

7. Real numbers!

Some non-examples

1. All matrices, not just those of the same size. They can be multiplied by scalars, but can't *always* be added.
2. Functions that have a limit at a point x_o. They can be multiplied by scalars, but can't *always* be added. The problem is like that for matrices of different sizes: two functions that have a limit at 0, for example, might not be defined on the same open interval about 0, so they can't necessarily be added pointwise. There is a way to get around this difficulty, but that is beyond the scope of a first course in linear algebra.
3. Matrices of the same size, with integer entries. These can be added, but can't be multiplied by scalars that are not integers, and still be matrices with integer entries.
4. Rational numbers. These can be added, but can't be multiplied by scalars that are not rational numbers, and have the results still be rational numbers.
5. Positive real numbers. These can be added, but can't be multiplied by scalars that are negative numbers, and have the results still be positive numbers.

Properties of the linear operations

When we work with mathematical objects, we usually take for granted that the following properties of addition and multiplication by constants - scalars, that is - are true. Therefore, in the theory they become a *requirement* that a mathematical object must satisfy in order that it be qualified to be called a linear space (a synonym for “vector space”).

0. There is an element of V, called zero, or (the) additive identity, and denoted by 0 (some times by $\mathbf{0}$ for emphasis, or to avoid confusion with the scalar 0), such that, for every vector v in V, $v + \mathbf{0} = v$.

1. (Vector) addition is commutative: for every v_1 and v_2 in V, $v_1 + v_2 = v_2 + v_1$.

2. Addition is associative: for every v_1, v_2, and v_3 in V, $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$.

3. Existence of additive inverses: for every v in V, there exists an element w of V such that $v + w = \mathbf{0}$. This vector w, called the additive inverse of v, is denoted by $-v$.

4. Scalar multiplication is associative: for all real numbers c_1 and c_2 and for every v in V, $c_1(c_2v) = (c_1c_2)v$.

5. Scalar multiplication is distributive over scalar addition: for all real numbers c_1 and c_2 and for every v in V, $(c_1 + c_2)v = c_1v + c_2v$.

6. Scalar multiplication is distributive over vector addition: for all real numbers c, and for every v_1 and v_2 in V, $c(v_1 + v_2) = cv_1 + cv_2$.

7. Multiplicative identity: for every v in V, $1v = v$.

These properties, together with the existence of the set V and the operations of addition and scalar multiplication that must always be performable, and result in an element of V, are required in order that we have a vector space. They are called the *axioms* of a vector space. Remember that the axioms include the stuff about the set V and the operations, even though that stuff did not appear in a numbered list! In the abstract theory, we simply *assume* the axioms are true, and can prove many theorems about objects defined in terms of the quantities in the axioms. Then, in an actual situation, if we can verify that what we're working with is a vector space, we then have available, for application, all the theorems that we have proved in the abstract setting. This is beneficial, because we do not have to verify over and over again, in different contexts, that what we do, involving only the linear operations, is correct. Other stuff we do, involving non-linear features of what we're working with, we can't apply the abstract theory to!

Exercise

Show that, in a vector space, for every scalar c, $c\mathbf{0} = \mathbf{0}$.

We have already seen some repetition, in the examples, of what you have to do to show something is a vector space. A labor-saving device has been invented - it's called a vector subspace, or linear subspace, and there is a theorem about them that can be used to save work.

Definition: Let V be a vector space. A subset, W, of V is called a <u>linear subspace</u> (or <u>vector subspace</u>) of V, if W, together with the same operations used in V, but applied only in W, is also a vector space.

Subspace Theorem: A subset, W, of a vector space V is a linear subspace of V if and only if for every pair w_1, w_2 of elements of W, and every pair c_1, c_2 of scalars, $c_1w_1 + c_2w_2 \in W$.

Exercise: Prove the theorem, by checking that all the axioms for a vector space are true for W, using the operations *already* OK for V.

This theorem is useful because, for example we can do the work of checking that the collection of real-valued functions defined on any non-empty set, with addition and scalar multiplication done pointwise, is a vector space. Then, to show that the functions that are differentiable on an interval form a vector space, all we have to do is show that $(af + bg)' = af' + bg'$ is true at every point - already done in calculus I - and apply the Subspace Theorem.

Exercise: Show that the examples (1 - 7) are indeed vector spaces.

Linear combination - a very useful term

Suppose $c_1, \dots, c_n$ are scalars, and $v_1, \dots, v_n$ are vectors in a vector space V. Then the vector $v = c_1v_1 + \dots + c_nv_n$ is a (the) linear combination of $v_1, \dots, v_n$ with coefficients $c_1, \dots, c_n$. A linear combination always involves a *finite* number of scalars and vectors! It might involve only one vector and one scalar. And the expression for v, namely $c_1v_1 + \dots + c_nv_n$, is written without parentheses (you see, in a legalistic sense, we can only add 2 vectors at a time, because that is what the operation "IZZ") because associativity lets us add them together in any order, 2 at a time. For example, $v_1 + v_2 + v_3 + v_4$ means $(v_1 + v_2) + (v_3 + v_4)$, or $((v_1 + v_2) + v_3) + v_4$, or (how many ways?) We also say that v is expressed as a linear combination of $v_1, \dots, v_n$, or represented as a linear combination of $v_1, \dots, v_n$. A linear combination is non-trivial if at least one of the scalars $c_1, \dots, c_n$ is non-zero. A linear combination is called trivial if all of the scalars $c_1, \dots, c_n$ in it are 0. Thus, a trivial linear combination always has a result, v, that is $\mathbf{0}$.

Linear dependence, linear independence, basis, and dimension

These 4 concepts are fundamental! Linear dependence and linear independence are mutually exclusive. They describe properties of *sets* of vectors, not properties of individual vectors! There is a tricky thing to remember about sets: a set can't have 2 copies of the same thing in it!

Definition: Let V be a vector space. A subset, S, of V is called a <u>linearly dependent</u> set if **0** is in S, or if there is a non-zero vector s in S that can be expressed as a linear combination of vectors in S, that are all *different* than s.

Alternate Definition: Let V be a vector space. A subset, S, of V is called a <u>linearly dependent</u> set if **0** (though not necessarily *in* S) can be expressed as a *non-trivial* linear combination of vectors in S.

Examples of linearly dependent *sets*:

1. S = {v, 2v, w}. If $v \neq \mathbf{0}$, then $v = 0w + \frac{1}{2}(2v)$. If $v = \mathbf{0}$, then 2v really is not there, because *sets* contain only one copy of each element. Then S is certainly linearly dependent by the first definition; it is linearly dependent by the second one because then $\mathbf{0} = 1\cdot\mathbf{0}$.

2. Suppose V is the set of polynomials of degree $\leq$ 2, and $S = \{ \mathbf{1}, \mathbf{x}, \mathbf{x}^2, (\mathbf{x} + \mathbf{1})^2 \}$. then S is a linearly dependent set because $(\mathbf{x} + 1)^2 = 1\cdot\mathbf{x}^2 + 2\cdot\mathbf{x} + 1\cdot\mathbf{1}$. We can express **0** as a non-trivial linear combination of the elements of S by re-writing the last equation in the form

$$\mathbf{0} = 1\cdot\mathbf{x}^2 + 2\cdot\mathbf{x} + 1\cdot\mathbf{1} + (-1)\cdot(\mathbf{x} + 1)^2.$$

3. Suppose you start with the matrix $\begin{pmatrix} 2 & 4 & 6 & 3 & 45 \\ 2 & 4 & 5 & 2 & 36 \\ 1 & 2 & 3 & 2 & 25 \\ 1 & 2 & 4 & 2 & 29 \end{pmatrix}$, and do Gauss elimination, getting the row-echelon matrix, $\begin{pmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. You can say, about the original matrix, that its rows form a linearly dependent set, because row 4 is a linear combination of rows 1, 2, and 3.

Exercise: Express row 4 of this matrix as a linear combination of rows 1, 2, and 3.

Exercise: Express sinh x and cosh x as linear combinations of e^x and e^{-x}.

Exercise: Express e^x and e^{-x} as linear combinations of sinh x and cosh x.

The definition of linearly independent is almost too easy: A set S in a vector space is linearly independent if it is *not* linearly dependent. This is correct, but not useful.

Alternate Definition: Let V be a vector space. A subset, S, of V is called a linearly independent set if $\mathbf{0}$ (necessarily not *in* S) can only be expressed as the *trivial* linear combination of vectors in S.

Alternate Alternate Definition: A set S in a vector space is linearly independent if the *only* sets $c_1, \dots, c_n$ of scalars that can make *even one* equation of the form $\mathbf{0} = c_1v_1 + \dots + c_nv_n$ true, whenever each of the vectors v_n is in S, are the sets with all of the c_i's equal to 0.

Examples:

1. $\{ \sin x, \cos x \}$ is a linearly independent set of functions. *Proof*: Suppose $\mathbf{a} \sin x + \mathbf{b} \cos x = \mathbf{0}$. In this context, $\mathbf{0}$ is the function that is identically zero. Therefore, if we define a function $f(x) = \mathbf{a} \sin x + \mathbf{b} \cos x$, we're saying that $f(x)$ is identically 0. But then, we can set $x = 0$, and we get $0 = f(0) = \mathbf{a}\cdot 0 + \mathbf{b}\cdot 1 = \mathbf{b}$, so $\mathbf{b} = 0$. Therefore, $f(x) = \mathbf{a} \sin x$ at every x. But $f(x) = 0$ at every x too. So choose some x where $\sin x \neq 0$. then $0 = f(x) = \mathbf{a} \sin x$, and this means that $\mathbf{a}$ has to be 0. Therefore the only constants $\mathbf{a}$ and $\mathbf{b}$ that make the equation $\mathbf{a} \sin x + \mathbf{b} \cos x \equiv 0$ true are $\mathbf{a} = 0 = \mathbf{b}$. Therefore, by the Alternate Alternate Definition, $\{\sin x, \cos x\}$ is a linearly independent set of functions.

2. The functions $\{1, x, x^2, \dots, x^n, \dots\}$ form a linearly independent set of functions. *Proof*: Suppose a linear combination of $1, x, x^2, \dots, x^n, \dots$ is identically zero. Since linear combinations only involve a finite number of these powers of x, the linear combination is a polynomial of some degree, say N. That is, we are given that $P(x) = p_o + p_1x + p_2x^2 + p_2x^2 + \dots + p_Nx^N \equiv 0$. Set x equal to 0. Since $0 = P(0) = p_o$, we get that the constant term is 0. Since $P(x) \equiv 0$, $P'(x) \equiv 0$. Again, set $x = 0$. Then $0 = P'(0) = p_1$. Repeat the

argument, and you'll get, at each stage, that $k!\, p_k = 0$. Hence, all the coefficients in the linear combination are 0, so the whole set is linearly independent.

3. Let V be the set of $n \times 1$ matrices - column vectors with n entries, (sometimes called components). Let S be the set consisting of the vectors $e_1, \ldots, e_n$, where e_i has all zero entries except for one 1 in the i^{th} row. Then S is a linearly independent set. *Proof*: Suppose $c_1e_1 + \ldots + c_ne_n = \mathbf{0}$. In this context, $\mathbf{0}$ is the column vector that has all entries 0. Now, the first entry in each of $e_2, \ldots, e_n$ is 0. Thus the first entry in $c_1e_1 + \ldots + c_ne_n$ is c_1. This is equal to the first entry in the sum, which is the first entry in $\mathbf{0}$, which is 0. That is, $c_1 = 0$. The argument just done works for *any* of the entries, so *all* the c_i's are 0.

Exercises:

1. Show that the columns of the matrix $\begin{pmatrix} 1 & 3 & 4 \\ 3 & 2 & 4 \\ 5 & 4 & 3 \end{pmatrix}$ form a linearly independent set.

2. Show that the rows of the matrix in Exercise 1 form a linearly independent set.

3. Show that the set $\{ e^x, e^{-x} \}$ is a linearly independent set of functions.

4. Show that any non-empty subset of a linearly independent set is a linearly independent set.

Theorems about linearly independent sets

The useful thing about linearly independent sets is that the coefficients used to represent a vector as a linear combination of the elements of a linearly independent set, if it can be done at all, can be done in only one way! Thus, if $v = c_1v_1 + \ldots + c_nv_n$, and $v = d_1v_1 + \ldots + d_nv_n$, then for every i, $c_i = d_i$. *Proof*:

$$\mathbf{0} = v - v = (c_1v_1 + \ldots + c_nv_n) - (d_1v_1 + \ldots + d_nv_n) = (c_1 - d_1)v_1 + \ldots + (c_n - d_n)v_n.$$

Since the v_i's are in a linearly independent set, by assumption, the coefficients, $(c_i - d_i)$ in the last equation must all be 0. That means $c_i = d_i$ for all the i's. Note - there is something wrong with this argument - nothing serious, mind you, but still not quite right! Can you spot the flaw?

Here is a theorem we will ***<u>use</u>***, but skip the proof of. You need to memorize it, and be able to use it. You can get a proof or explanation from me if you wish, later. In order to state the theorem we need the concept of a set that is maximal, in a set X, among those sets that have some property, say property **P**. For this to be useful, we will only talk about properties of sets that are <u>hereditary</u>, meaning that if a set S has propery **P**, and T is a subset of S, then T also has property **P**. An example of such a property of sets is the property that the set does not contain **0**. That's not an interesting example. The one we will use is the property of being linearly independent. It is actually true that the empty set is linearly independent! So Exercise 4, above, that I hope you did, shows that being linearly independent is an hereditary property. So, saying a set is <u>maximal</u> with respect to property **P** means that the set has property **P**, and that no larger set has property **P**. There might be many maximal sets.

Theorem: In every vector space V there exists at least one maximal linearly independent set B.

There are, in fact, infinitely many sets of this sort. But they all have one thing in common:

Theorem: If B is a maximal linearly independent set in a vector space V, then every vector, v, in V can be represented, in one and only one way, as a linear combination of the vectors in B.

Proof: Let $v \in V$. Then, if $v \in B$, $v = 1 \cdot v$, so v *can* be represented as a linear combination of the elements of B. Since B is linearly independent, this can be done in only one way. On the other hand, if v does not belong to B, consider the set $B \cup$*** Unmatched right merge quote *** $\{ v \}$. Since this set is larger than B, it is not linearly independent - it is linearly dependent. This means that there is a finite set of elements of $B \cup$*** Unmatched right merge quote *** $\{ v \}$, and a corresponding set of scalars, *not all zero*, such that the linear combination, formed from the vectors and their corresponding scalars, is $\mathbf{0}$. Now, that finite set of elements of $B \cup$*** Unmatched right merge quote *** $\{ v \}$ must contain v, for otherwise it would be a subset of B, and we know, since B is linearly independent, that the *only* sets $c_1, \ldots, c_n$ of scalars that can make *even one* equation of the form $\mathbf{0} = c_1v_1 + \ldots + c_nv_n$ true, whenever each of the vectors v_n is in B, are the sets with *all* of the c_i's equal to 0. And we know that *our* subset of $B \cup$*** Unmatched right merge quote *** $\{ v \}$ is linearly dependent. So, v has to be one of its members. Let's denote this set $\{ v, v_1, \ldots, v_k \}$, with corresponding scalars $c, c_1, \ldots, c_n$ *not all of which are zero,* such that the linear combination

$$cv + c_1v_1 + \ldots + c_nv_n = \mathbf{0}.$$

I assert that c cannot be 0. For, otherwise, $c = 0$, so one or more of the c_i's have to be non-zero, and we'd have an equation

$$c_1v_1 + \ldots + c_nv_n = \mathbf{0},$$

involving only vectors from B, with coefficients not all 0. This can't be, because B is linearly independent. Therefore, $c \neq 0$. So, now we're done, because we can solve for v:

$$v = -(1/c)(c_1v_1 + \ldots + c_nv_n).$$

The part of the theorem about uniqueness of the representation of a vector as a linear combination of the elements in B is a consequence of the Theorem about linearly independent sets, proved earlier in this section.

Note*: see if you can get the last equation from the one before it, just using the axioms of a vector space.*

Well, now that that's done, the rest is not so hard. The set B that we get from the big theorem is called a basis of, or, for, V, or just a basis. But it looks intimidating to deal with a basis, much less find one, or be able to show that a particular set is a basis. Actually, it's usually not a big problem, and that's because of the theorem following the next theorem. To state the next 2 theorems, we need yet another set of handy terms: span (verb), span (noun), and associated adjectives. A set S of vectors spans a set T of vectors, if every element of T is (i.e., can be expressed as) a linear combination of vectors in S. The span of a set S of vectors is, by definition, the set of *all* linear combinations of vectors in S. The span of S is denoted *span* S.

Theorem: The span of a non-empty set of vectors in a vector space V is a linear subspace of V.

Proof: Let W denote the span of S, namely the set of all linear combinations of elements of S. By the subspace theorem, all we have to do is show that every linear combination of 2 elements of W is also in W. So, let $w_1 = c_1s_{1,1} + \ldots + c_ns_{1,n}$, $w_2 = d_1s_{2,1} + \ldots + d_ms_{2,m}$, where the double subscripts indicate that the vectors involved in forming w_1 and w_2 might not be the same. Then, for any scalars a_1 and a_2,

$$a_1w_1 + a_2w_2 = c_1s_{1,1} + \ldots + c_ns_{1,n} + d_1s_{2,1} + \ldots + d_ms_{2,m}.$$

This is a linear combination of vectors in V, except that some vectors might be repeated. That is actually OK, but we can eliminate repetitions by using the commutativity and associativity of addition, and the distributivity of scalar addition over scalar multiplication, all found in the axioms, to express $a_1w_1 + a_2w_2$ as a linear combination of the vectors in the union of the two sets

$\{s_{1,1}, \ldots, s_{1,n}\}$ and $\{s_{2,1}, \ldots, s_{2,m}\}$. This completes the proof of this handy theorem.

Now the theorem that makes bases (plural of basis) more or less easy to find.

Theorem: A subset B of a vector space V is a basis of V if and only if

(1) V = *span* B

and

(2) B is a linearly independent set.

Proof: (1) says, every vector in V is expressible as a linear combination of the vectors in B, and (2) says that this can be done in only one way. Therefore, the conditions to be a basis are satisfied by any B that satisfies the conditions (1) and (2). This completes the proof of this handy theorem.

Example: The set $\{1, x, x^2, \ldots, x^n, \ldots\}$ is a basis of the set of all polynomials. *Proof*: Every polynomial can be expressed as a linear combination of these powers of x. In one of the earlier examples, it was shown that this set of functions is a linearly independent set. Thus, the last theorem says the set of powers of x is a basis of the set of all polynomials.

Exercises:

1. Let $\mathbf{R}^n$ denote the set of $n\times 1$ matrices - column vectors with n entries, (sometimes called <u>components</u>). Let S be the set consisting of the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, where $\mathbf{e}_i$ has all zero entries, except for a 1 in the i^{th} row. Show that S is a basis for $\mathbf{R}^n$. (It's called the <u>standard basis</u> of $\mathbf{R}^n$.)

2. Let B denote the set of vectors in $\mathbf{R}^n$ of the form $b_i = e_i + e_{i+1}$, if $i < n$, and let $b_n = e_n$. Show that B is a basis for $\mathbf{R}^n$.

Next, we define, and deal with, the idea of dimension.

Definition: If V is a vector space, and B has a basis that is finite, then V is called a <u>finite-dimensional vector space</u>. If every basis of V is an infinite set, then V is called an <u>infinite-dimensional vector space</u>.

Theorem and Definition: If a vector space V has one infinite basis, then each of its bases is infinite. If a vector space is finite dimensional, then each of its bases has the same number of elements, and that number is called the <u>dimension</u> of V.

Proof: Suppose that B_1 and B_2 are bases of V. Suppose that B_1 is an infinite set, and that B_2 is a finite set. We will show that this leads to a contradiction. The idea of the proof is to use systems of linear equations! Suppose that B_2 has N elements; say $B_2 = \{s_{2,1}, \dots, s_{2,N}\}$. Choose $N+1$ elements from B_1. This can be done, because B_1 is an infinite set. Call the chosen elements $s_{1,1}, \dots, s_{1,N+1}$. Since B_2 is a basis, for each k, $1 \le k \le N+1$, there exist scalars c_{kj} such that $s_{1,k} = \sum_{j=1}^{N} c_{kj}s_{2,j}$. Now, given *any* set of $N+1$ x_k's, let's form the linear combination $w = \sum_{k=1}^{N+1} x_k s_{1,k}$, and substitute in, for each of the $s_{1,k}$, the expression for $s_{1,k}$, in terms of B_2. This gives the complicated expression $w = \sum_{k=1}^{N+1} x_k s_{1,k} = \sum_{k=1}^{N+1} x_k \sum_{j=1}^{N} c_{kj}s_{2,j}$. By using the usual properties of addition and scalar multiplication, we can change the order of summation, and get

$$\sum_{k=1}^{N+1} x_k s_{1,k} = \sum_{j=1}^{N} \left(\sum_{k=1}^{N+1} x_k c_{kj} \right) s_{2,j} = w.$$

Of course, we knew all along, because B_2 is a basis, that there are *unique* numbers y_j such that $w = \sum_{j=1}^{N} y_j s_{2,j}$. Therefore, since $\sum_{j=1}^{N} \left(\sum_{k=1}^{N+1} x_k c_{kj} \right) s_{2,j} = w = \sum_{j=1}^{N} y_j s_{2,j}$, we know, for each j, $1 \le j \le N$, the coefficients of $s_{2,j}$ on each side of the equation must be the same, so $y_j = \sum_{k=1}^{N+1} x_k c_{kj}$. This is a system of linear equations! In fact, there are N equations in $N+1$ unknowns. Therefore, if there are solutions at all, there must be infinitely many. But there *are* solutions, because we started with any set of $N+1$ x_k's, and built a vector, w, as a linear combination of $N+1$ vectors chosen from B_1. Then we sent out for the coefficients of w with respect to the basis B_2. These were the y_j's, N of them. We found 2 ways to represent the vector w in terms of the B_1, one way as a sum involving x_k's, and the other by just asking for the coefficients of w with respect to B_2. Because

B_2 is linearly independent, the 2 ways can't really be different. That's how we got the equations $y_j = \sum_{k=1}^{N+1} x_k c_{kj}$, $1 \le j \le N$. So there certainly are solutions! Since there have to be others, we also have numbers z_k, $1 \le j \le N+1$, not all the same as the corresponding x_k's, such that $y_j = \sum_{k=1}^{N+1} z_k c_{kj}$, for $1 \le j \le N$. Now we have to work backwards. Certainly $w = \sum_{j=1}^{N} y_j s_{2,j} = \sum_{j=1}^{N} \sum_{k=1}^{N+1} z_k c_{kj}\, s_{2,j}$. Now let's change the order of summation. We get $w = \sum_{j=1}^{N} \sum_{k=1}^{N+1} z_k c_{kj}\, s_{2,j} = \sum_{k=1}^{N+1} z_k s_{1,k}$.

The last equality comes from substituting "out" the representation of the $s_{1,k}$'s in terms of B_2. Now we are done, because we have the result that $w = \sum_{k=1}^{N+1} z_k s_{1,k}$, even though we know that $w = \sum_{k=1}^{N+1} x_k s_{1,k}$, and that some of the x_k's differ from the corresponding z_k's, so this contradicts the linear independence of B_1: corresponding coefficients *must* be the same when the vectors lie in a linearly independent set!

What we have shown so far is that, if one basis is finite, so must be every other one. This is the same thing as saying that, if one basis of V is *not* finite, then all bases of V are infinite. It remains to show that, when one basis of V is finite, and contains n elements, then every basis of V contains exactly n elements. Well, we can recycle the previous argument - it already deals with this case too. Here's how: Suppose not. Then there are bases of different sizes. Call the larger one B_1, the smaller, B_2. Let N denote the number of elements in B_2. Now use the previous argument, almost word for word. The only difference is, we can choose N+1 elements from B_1, not because it is infinite, but because it has more elements than B_2. Now, since every basis of V has the same number of elements, we can legitimately say that that number depends on the vector space itself, and not on any particular basis of it. This number is called the <u>dimension</u> of V.

Exercise: Explain why a subspace of a finite dimensional vector space is itself a finite dimensional vector space.

Linear transformations

Linear transformations are where the applications are, of linear algebra. We'll begin with the definition, and then list some examples.

Definition: A function T that is defined on a vector space V, and that takes its values in a vector space is called a linear transformation, or a linear map(ping), or simply linear, if it is true that, for every pair v_1, v_2 of elements of V, and every pair c_1, c_2 of scalars,

$$T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2).$$

"T linear if it preserves linear combinations."

The notation $T: V \to W$ is frequently used in describing linear transformations - it emphasizes the domain (space), V, of T, and the range-space of T, namely W.

Exercise: Show that T is linear if and only if, for every pair v_1, v_2 of elements of V, $T(v_1 + v_2) = T(v_1) + T(v_2)$, and, for every v in V and every scalar c, $T(cv) = cT(v)$. That is, "T is linear if it preserves sums and scalar multiplication."

Examples:

1. Let A be an $m\times n$ matrix. Then define $T(x)$, for column vectors of size n, by $T(x) = Ax$. $T(x)$ is a column vector of size m. The basic properties of the matrix operations tell us that $T(x_1 + x_2) = A(x_1 + x_2) = Ax_1 + Ax_2 = T(x_1) + T(x_2)$. Also, $T(cx) = A(cx) = (Ac)x = (cA)x = c(Ax) = cT(x)$.

2. Let V be the set of functions that are differentiable at every point of the unit interval. For f in V, let $T(f)$ be the function of x defined by $f'(x)$. Thus, $(T(f))(x) = f'(x)$. Since

$$(f + g)'(x) = f'(x) + g'(x),$$

and

$$(cf)'(x) = c(f'(x)),$$

we see, by recognizing the occurences of the definition of T, that T is linear.

Exercise: Show that the function that starts with a continuous function f on [0,1], and assigns to it the number $\int_0^1 f(x)\,dx$, is a linear transformation. This is an example of a linear transformation that has scalars as values. Such are called <u>linear functionals</u>. They are a very important kind of linear transformation.

Usually, we write a linear transformation not as T(x), but as Tx, to emphasize the analogy with multiplication by a matrix.

Caution: The word "linear" is used in different ways in mathematics. You are probably used to calling a polynomial such as mx + b a linear function, and that makes sense, because its graph is a line. But, if b is not 0, it is not a linear function *in the sense of linear algebra*. *Proof*: In order for T(x) = mx + b to be linear in the sense of linear algebra, it has to be true that for every scalar c, T(cx) = cT(x). Let c = 0. Then T(cx) = T(0) = b. But cT(x) = 0T(x) = 0 $\neq$ b, so T is not linear. See if you can prove T is not linear using addition instead of scalar multiplication.

Exercise: Show that, if T is linear, then T(**0**) = **0**.

Here are some terms that you'll need to know later, about a linear map T: V $\rightarrow$ W:

The <u>null space</u>, or <u>kernel</u>, of a linear transformation is denoted *ker* T, and is defined by

$$\ker T = \{ v \text{ in } V : T(v) = \mathbf{0} \}.$$

Theorem: The null space of a linear map T is a subspace of V, the vector space where T is defined.

The <u>range</u>, or <u>image</u> of a linear transformation is denoted *im* T, and is defined by

$$\operatorname{im} T = \{ w \text{ in } W : \text{for some } v \text{ in } V,\ w = T(v) \}.$$

Theorem: The range of a linear map T is a subspace of W, the vector space where T takes its values.

The solutions X of a homogeneous matrix equation AX = 0 form its null space. The column operation matrix C that is mentioned at the end of section 5 contains, in its n-r columns under the all-zero columns of the "final cleaned up" matrix, a basis for the null space.

The span of the columns of A is the range of A. A basis for the range of A can be found by multiplying A by the matrix formed by the columns of C that are under the $r \times r$ identity matrix that fills the first r columns of the "final cleaned up" matrix. The product has r columns, each of size m, and they form a basis for the range of A.

How the dot product is related to a sum of products

The geometric definition of the dot product of vectors **F** and **r** in space is $\mathbf{F} \cdot \mathbf{r} = ||\mathbf{F}||\ ||\mathbf{r}|| \cos\theta$, where θ is the angle between **F** and **r**. We measure θ so that it is between 0 and π, inclusive.

The matrix, or coordinate, definition of the dot product: If x and y are vectors in $\mathbf{R}^n$, then $x \cdot y = \sum_{i=1}^{n} x_i y_i$. The geometric definition in $\mathbf{R}^n$ depends on the fact that the span of a set with 2 elements has dimension at most 2. Then "the angle between x and y" makes sense. So we define geometric $x \cdot y = ||x||\ ||y|| \cos\theta$, where θ is the angle between x and y. We measure θ so that it is between 0 and π, inclusive.

The fact is that these definitions yield the same number. Why? It all depends on someting that is geometric - the Law of Cosines: $c^2 = a^2 + b^2 - 2ab\cos\theta$, where θ is the angle between the sides of the triangle that have lengths a and b, respectively, and c is the length of the opposite side. Please draw a picture! The way we bring vectors into the picture is to interpret these lengths as the magnitudes of vectors, and use Pythagoras to say that the square of the magnitude is the sum of the squares of the entries of the vector. So let $a = ||x||$, $b = ||y||$. What is c, in terms of x and y? Well, by the parallelogram rule of vector addition, $c = ||x - y||$ or $||y - x||$, depending on how you drew your picture. It will work either way! Substitute these into the Law of Cosines: $||x - y||^2 = ||x||^2 + ||y||^2 - 2||x||\ ||y|| \cos\theta$, and now put all the squares on the RHS, and the other term on the LHS: $2||x||\ ||y|| \cos\theta = ||x||^2 + ||y||^2 - ||x - y||^2 = \sum_{i=1}^{n} \{x_i^2 + y_i^2 .- (x_i - y_i)^2\} = \sum_{i=1}^{n} 2x_i y_i$. We can cancel the 2's, and the equation says the two definitions agree.

The next few subsections all lead to systems of linear equations, and make use of matrix operations, especially the inverse. It is important for you to be able to recognize a linear situation when you see it. These subsections illustrate that too.

Coordinate systems

Every vector space has a basis. If **V** is a finite dimensional vector space, and $V = \{ v_1, \dots, v_{\dim\mathbf{V}} \}$ is a basis of **V**, then every vector v in **V** can be written as a linear combination of the v_j's, with scalars c_j, called <u>coefficients</u>. For simplicity, let's let $n = \dim\mathbf{V}$ = dimension of **V**. Then $v = c_1 v_1 + \dots + c_n v_n$. We also know that the coefficients are *unique*, because the basis V is linearly independent. Why? Because, if $v = c_1 v_1 + \dots + c_n v_n$, and also $v = c_1' v_1 + \dots + c_n' v_n$, then $v - v = c_1 v_1 + \dots + c_n v_n - (c_1' v_1 + \dots + c_n' v_n)$

$= (c_1 - c_1')v_1 + \ldots (c_n - c_n')v_n = 0$. Since a basis is a linearly independent set, all the coefficients $(c_i - c_i') = 0$, that is, $c_i = c_i'$ for $i = 1, \ldots, n$. Thus, there is only one set of coefficients for v: the <u>coefficients are unique</u>. The discussion so far has been, and will continue to be, abstract - nothing is being said about HOW those coefficients are found! That depends on the exact nature of each particular vector space. Right now the discussion is about things that are true for vector spaces in general. Now for coordinates. Let $C(v) = (c_1, \ldots, c_n) = (c_1 \ldots c_n)^t$. I claim that C is a linear transformation from **V** to $\mathbf{R}^n$. To show this, I can use the two-step method: show that, for all vectors v, w in **V**, $C(v + w) = C(v) + C(w)$, and for all scalars r, and all vectors v, $C(rv) = rC(v)$. Let's suppose that $w = d_1v_1 + \ldots + d_nv_n$, so that the coefficients of w are $d_1, \ldots, d_n$. Therefore, $C(w) = (d_1, \ldots, d_n)$. Hence $C(v) + C(w) = (c_1, \ldots, c_n) + (d_1, \ldots, d_n) = (c_1 + d_1, \ldots, c_n + d_n)$. Does this equal $C(v + w)$? Well, $v = c_1v_1 + \ldots + c_nv_n$, and $w = d_1v_1 + \ldots + d_nv_n$, so $v + w = (c_1 + d_1)v_1 + \ldots + (c_n + d_n)v_n$, so the coefficients of $v + w$, being unique because the v_j's constitute a basis, are $(c_1 + d_1), \ldots, (c_n + d_n)$, so $C(v + w) = ((c_1 + d_1), \ldots, (c_n + d_n)) = C(v) + C(w)$. The argument for $C(rv) = rC(v)$ is similar. You do it, please! We call the column vectors C(v) the <u>coordinates of</u> v, or the <u>coordinate vector</u> of v, <u>with respect to the basis</u> V. So the coordinates we use to specify a vector depend on the basis we use. The vectors in the basis are sometimes called a <u>coordinate frame</u>. If we want to keep track of the coordinate frame we are using, we can write $C_V(v)$

The coordinate transformation can be reversed. If we start with a vector $(c_1, \ldots, c_n)$ in $\mathbf{R}^n$, and use its entries to form the vector $v = c_1v_1 + \ldots + c_nv_n$ in **V**, then $C(v) = (c_1, \ldots, c_n)$.

The rule just given that assigned $c_1v_1 + \ldots + c_nv_n$ in **V** to $(c_1, \ldots, c_n)$ in $\mathbf{R}^n$ defines a function

$K((c_1, \ldots, c_n))$ that maps $\mathbf{R}^n$ to **V**.

Exercise: Show that K is a linear transformation from $\mathbf{R}^n$ to **V**.

I claim that K is the inverse of C. To show why this is true I need to check the defining equations for inverses: $KC = I$ (the identity on **V**) and $CK = I$ (the identity on $\mathbf{R}^n$). From the way K was defined, it has already been shown that $CK = I$ (on $\mathbf{R}^n$). Right? Now I need to check that $KC = I$ (on **V**). So start with v in **V**. C(v) is the vector of coordinates of v. That is, $(c_1, \ldots, c_n)$ is that vector of scalars such that $v = c_1v_1 + \ldots + c_nv_n$. And so $KC(v) = c_1v_1 + \ldots + c_nv_n$. But this is v. So $KC = I$ (on **V**). We will use C^{-1} to denote the inverse, not K, or we'll use C_V^{-1} if we need to emphasize that we're working with the basis V.

Example: The vectors $\{ e_1, \ldots, e_n \}$ in $\mathbf{R}^n$ form a basis, called the <u>standard basis</u>, and $C(v) = v$.

Example: Now let's work with the vectors (3, 4) and (4, 3) in $\mathbf{R}^2$. They form a basis (why?). Let's see how to find out what $C(\begin{pmatrix} x \\ y \end{pmatrix})$ is. We need to find a and b so that $\begin{pmatrix} x \\ y \end{pmatrix} = a\begin{pmatrix} 3 \\ 4 \end{pmatrix} + b\begin{pmatrix} 4 \\ 3 \end{pmatrix}$, and a and b will depend on x and y AND the vectors in the basis, AND the order of the vectors in the basis. Let's look at what's happening at the entry level: $x = 3a + 4b$, and $y = 4a + 3b$. This is a system of 2 equations in 2 unknowns, that can be expressed in matrix form as $\begin{pmatrix} 3 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$. To solve for arbitrary x and y given on the right, we find the inverse of $\begin{pmatrix} 3 & 4 \\ 4 & 3 \end{pmatrix}$, which is

$-\frac{1}{7}\begin{pmatrix} 3 & -4 \\ -4 & 3 \end{pmatrix}$, and find that $\begin{pmatrix} a \\ b \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} 3 & -4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$. Thus, since a and b are the coordinates desired, $C(\begin{pmatrix} x \\ y \end{pmatrix}) = -\frac{1}{7}\begin{pmatrix} 3x - 4y \\ -4x + 3y \end{pmatrix}$. Notice that the coordinates of $\begin{pmatrix} x \\ y \end{pmatrix}$ with respect to

$V = \{ \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix} \}$ are NOT x and y.

Exercise: What would the coordinates of $\begin{pmatrix} x \\ y \end{pmatrix}$ be with respect to $W = \{ \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix} \}$?

Example: Now let **V** be the space of polynomials of degree ≤ 3. A standard basis is $\{1, x, x^2, x^3\}$. The coordinates of $ax^3 + bx^2 + cx + d$ are therefore (d, c, b, a). Surprised? The reason for the reversal of order is that we listed the vectors in the basis in increasing order of powers. The polynomials $1, 1+x, (1+x)^2, (1+x)^3$ also form a basis (why??). What is $C(ax^3 + bx^2 + cx + d)$ for this new basis? To find out, we have to figure out how to express $ax^3 + bx^2 + cx + d$ as a linear combination of $\{ 1, 1+x, (1+x)^2, (1+x)^3 \}$. So, we need to solve for A, B, C, D so that $ax^3 + bx^2 + cx + d = A1 + B(1+x) + C(1+x)^2 + D(1+x)^3$. We could work from the right-hand side by choosing various values of x, and see what that yields. Choose $x = -1$. The equation becomes $-a + b - c + d = A$. So now we know what A, the first coordinate, is. We might next choose $x = 0$, and this gives the equation $d = A + B + C + D$, or $a - b + c = B + C + D$. Next, we could choose $x = -2$, and get $-8a + 4b - 2c + d - A = -B + C - D$, or $-7a + 3b - c = -B + C - D$. Finally, choose $x = 1$, and get $a + b + c + d = A + 2B + 4C + 8D$, or $2a + 2c = 2B + 4C + 8D$. This gives us a system of equations in the unknowns B, C, and D, and “knowns” a, b, and c (notice that d disappeared):

$$B + C + D = a - b + c,$$
$$-B + C - D = -7a + 3b - c,$$
$$2B + 4C + 8D = 2a + 2c.$$

This has the matrix form $\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 2 & 4 & 8 \end{pmatrix}\begin{pmatrix} B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 3 & -7 \\ 2 & 0 & 2 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix}$. Notice the reverse order of c, b, a! Remember, that’s because of the choice of order made in the basis! The 3×3 matrix on the left has a special form! Its first column consists of different numbers. Its subsequent columns are powers, second, third, and so on (such matrices can be of any size) of the numbers in the first column. This is a <u>Vandermonde matrix</u>, and it is invertible. I do not know a simple way to invert a Vandermonde matrix - maybe there is an article in a journal in the Math Library about it, though. The inverse of this one is

$\frac{1}{6}\begin{pmatrix} 6 & -2 & -1 \\ 3 & 3 & 0 \\ -3 & -1 & 1 \end{pmatrix}$. Therefore, $\begin{pmatrix} B \\ C \\ D \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 6 & -2 & -1 \\ 3 & 3 & 0 \\ -3 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 3 & -7 \\ 2 & 0 & 2 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix} = \begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} c \\ b \\ a \end{pmatrix}$. This can be combined, using block matrix techniques, with the formula for A, $A = d - c + b - a$, to

get the coordinates: $\begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} d \\ c \\ b \\ a \end{pmatrix}$. Well, this is a way that works, and it’s very matricial (aha! the adjective at last!). But there is an easier way! In the original polynomial, write $x = ((x+1) - 1)$, and plug this expression in for x: $ax^3 + bx^2 + cx + d = a((x+1) - 1)^3 + b((x+1) - 1)^2 + c((x+1) - 1) + d =$

$$\begin{aligned} &a(x+1)^3 - 3a(x+1)^2 + 3a(x+1) - a \\ &\quad + b(x+1)^2 - 2b(x+1) + b \\ &\quad\quad + c(x+1) - c \\ &\quad\quad\quad + d = \end{aligned}$$

$(d - c + b - a)1 + (c - 2b + 3a)(x+1) + (b - 3a)(x+1)^2 + a(x+1)^3$. This is exactly what the matix approach gave as the answer.

Exercise: Assume Earth is spherical for this exercise. In terms of a right-handed coordinate system with origin in Minneapolis at the spot where the 45° line of North lattitude meets the $93^\circ\ 15'$ line of West longitude, with UP as the z-direction, EAST as the x-direction, (a) find the coordinates of the center of the Earth (assume the radius is 3500 miles), (b) find the coordinates of the Greenwich Observatory. You'll need to consult an Atlas for (b)!

Conclusion: Finding the coordinates of a vector with respect to a basis depends on the nature of that basis, and the nature of the vector space in which all this takes place. The easiest basis to work with is some kind of natural one, such as the standard basis in $\mathbf{R}^n$, and the powers of x in spaces of polynomials. But sometimes, it is necessary to use some other basis in place of the easy one, so we have to know about <u>change of basis</u>. Note that in the examples, the vectors to find the coordinates of were given in terms of a standard basis!

Change of basis

Suppose that $\mathbf{V}$ is a finite-dimensional vector space, of dimension $n > 1$. Assume we have two bases, $V = \{ v_1, \dots , v_n \}$ (so maybe V is a standard basis of $\mathbf{V}$) and $W = \{ w_1, \dots , w_n \}$. We know (abstractly) that there are coordinates, $C_V(v)$ and $C_W(v)$ for each vector v in $\mathbf{V}$, with respect to each of these bases. If we are given $C_V(v)$, the coordinates of v with respect to V, is there a way to find $C_W(v)$, the coordinates of v with respect to W, in terms of $C_V(v)$? Yes. We "construct" (i.e., find, in principle, or in an abstract way) an $n\times n$ <u>change-of-basis matrix</u>, X, (for eXchange) such that $C_W(v) = XC_V(v)$ (recall that $C_V(v)$ and $C_W(v)$ are in $\mathbf{R}^n$, so they're $n\times 1$ column vectors). Here's how. We're given that $v = \sum_{j=1}^{n} c_jv_j$, and we know that there exist numbers d_i such that $v = \sum_{i=1}^{n} d_iw_i$, but we don't know them. Now, somehow, I can't say how unless I know what the vector space $\mathbf{V}$ actually is, we dig around and find, for each j, $1 \le j \le n$, unique numbers b_{ij}, such that $v_j = \sum_{i=1}^{n} b_{ij}w_i$, because $\{ w_1, \dots , w_n \}$ is a basis too. So here we are, scurrying around, looking up the coordinates for n vectors, when all we wanted was the coordinates for one vector! Well, not exactly - the vector we want the w-coordinates for is an *arbitrary* vector, v, and there are infinitely many of them. If we just do this work of finding the w-coordinates of n vectors, we can find the w-coordinates of *any* vector, just by doing a matrix multiplication.times the v-coordinates of that same vector! Of course, finding the v-coordinates might be hard. So we only have to do the hard thing once… Well, now we plug the expressions $v_j = \sum_{i=1}^{n} b_{ij}w_i$ into the equation $v = \sum_{j=1}^{n} c_jv_j$, and we get a double sum $v = \sum_{j=1}^{n} c_j \sum_{i=1}^{n} b_{ij}w_i = \sum_{i=1}^{n}\sum_{j=1}^{n} c_jb_{ij}w_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} c_jb_{ij}\right)w_i$. We also know (in principle) that $v =$

$\sum_{i=1}^{n} d_i w_i$, so if we subtract these two expressions for v, we get

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n} c_j b_{ij}\right) w_i - \sum_{i=1}^{n} d_i w_i = \sum_{i=1}^{n}\left(\left(\sum_{j=1}^{n} c_j b_{ij}\right) - d_i\right) w_i = 0.$$

Since $\{ w_1, \dots , w_n \}$ is a linearly independent set, we know that the numbers $\left(\sum_{j=1}^{n} c_j b_{ij}\right) - d_i$ must all be 0, $i = 1, \dots , n$. Therefore, if we let $X_{ij} = b_{ij}$, then $\sum_{j=1}^{n} X_{ij} c_j = \sum_{j=1}^{n} c_j b_{ij} = d_i$, for $i = 1, \dots , n$. So, $C_W(v) = (d_1, \dots ,d_n) = X(c_1, \dots ,c_n) = XC_V(v)$, as desired. The matrix X is called the change-of-basis matrix, from V to W. We can write $X_{V,W}$ for X if we need to keep track of the bases we're exchanging. The order of the subscripts V and W is important!

Exercise: Figure out, just by thinking about it, how now to find the change-of-basis matrix from W to V, assuming you already know X.

Example: Find the change-of-basis matrix from $\left\{ \begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}4\\5\\6\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix} \right\}$ to $\left\{ \begin{pmatrix}2\\-1\\0\end{pmatrix}, \begin{pmatrix}-1\\2\\-1\end{pmatrix}, \begin{pmatrix}0\\-1\\2\end{pmatrix} \right\}$. We have 3 sets of coefficients to find. First, find b_{11}, b_{12}, b_{13} so that $\begin{pmatrix}1\\2\\3\end{pmatrix} = b_{11}\begin{pmatrix}2\\-1\\0\end{pmatrix} + b_{21}\begin{pmatrix}-1\\2\\-1\end{pmatrix} + b_{31}\begin{pmatrix}0\\-1\\2\end{pmatrix}$. Please recognize this equation as one with a matrix product: $\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}\begin{pmatrix}b_{11}\\b_{21}\\b_{31}\end{pmatrix}$. Now recognize that when we do the same thing for the other two basis elements, we can recognize a matrix equation (this is possible because we are working with column vectors!):

$\begin{pmatrix}1&4&1\\2&5&2\\3&6&1\end{pmatrix} = \begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}\begin{pmatrix}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{pmatrix}$. This has the form $V = WX$, so $X = W^{-1}V$. I used V to stand for the matrix consisting of the column vectors in V, likewise for W. I can do so ONLY because I'm working in $\mathbf{R}^3$! So what we need to DO to find X is, find the inverse of W, the matrix formed from the vectors in the second basis. The inverse of W turns out to be $\frac{1}{4}\begin{pmatrix}3&2&1\\2&4&2\\1&2&3\end{pmatrix}$. Finally, we multiply on the right by V to get X. The answer is: $X = \frac{1}{2}\begin{pmatrix}5&14&4\\8&20&6\\7&16&4\end{pmatrix}$. Please check it! Let's see how this works. Suppose we know the coordinates of some vector v with respect to the basis V. That is, $v = c_1\begin{pmatrix}1\\2\\3\end{pmatrix} + c_2\begin{pmatrix}4\\5\\6\end{pmatrix} + c_3\begin{pmatrix}1\\2\\1\end{pmatrix} = Vc = WXc$. We also know, abstractly, that

$v = d_1\begin{pmatrix}2\\-1\\0\end{pmatrix} + d_2\begin{pmatrix}-1\\2\\-1\end{pmatrix} + d_3\begin{pmatrix}0\\-1\\2\end{pmatrix} = Wd$, but we don't know the d's. However, $WXc = Wd$, because each of these is equal to v. Since W is invertible, we can multiply on both sides by W^{-1}, and we get that $d = Xc$. But $c = C_V(v)$, and $d = C_W(v)$, so we've found the X that transforms V-coordinates into W-coordinates: $C_W(v) = XC_V(v)$.

Example(continued): Find $C_W(v)$. This means, given a vector v in $\mathbf{R}^3$, find d_1, d_2, d_3 such that $v = d_1\begin{pmatrix}2\\-1\\0\end{pmatrix} + d_2\begin{pmatrix}-1\\2\\-1\end{pmatrix} + d_3\begin{pmatrix}0\\-1\\2\end{pmatrix} = Wd$ (we recognize a linear combination of columns as a matrix times the column made of the coefficients in the linear combination!). To solve $v = Wd$ we multiply both sides by the inverse of W: $d = W^{-1}v$. We already know W^{-1}, so $C_W(v) = W^{-1}v$.

Exercise: Find $C_V(v)$. Find $C_V(v)$ when $v = (7, 11, 19)$.

Example(continued): Let $v = \begin{pmatrix}1\\2\\3\end{pmatrix} - 2\begin{pmatrix}4\\5\\6\end{pmatrix} + 5\begin{pmatrix}1\\2\\1\end{pmatrix} = \begin{pmatrix}-2\\2\\-4\end{pmatrix}$. Find the coordinates of v with respect to the basis W. The coefficients of v with respect to V are 1, –2, 5, so $C_V(v) = \begin{pmatrix}1\\-2\\5\end{pmatrix}$. Then $C_W(v) = XC_V(v) = X\begin{pmatrix}1\\-2\\5\end{pmatrix} = \frac{1}{2}\begin{pmatrix}5 & 14 & 4\\8 & 20 & 6\\7 & 16 & 4\end{pmatrix}\begin{pmatrix}1\\-2\\5\end{pmatrix} = -\frac{1}{2}\begin{pmatrix}3\\2\\5\end{pmatrix}$. Please check: is $\begin{pmatrix}-2\\2\\-4\end{pmatrix} = -\frac{3}{2}\begin{pmatrix}2\\-1\\0\end{pmatrix} - \begin{pmatrix}-1\\2\\-1\end{pmatrix} - \frac{5}{2}\begin{pmatrix}0\\-1\\2\end{pmatrix}$? Do the coordinates obtained this way agree with $W^{-1}v$?

Exercise: Find the coordinates of the standard basis vectors with respect to each of the bases in the last example. This exercise can be done with 3 matrix multiplications as the only labor, if you think through what is going on!

Exercise: Find the change-of-basis matrix from $\{ 1, 1+x, (1+x)^2 \}$ to $\{ 1, 1-x, (1-x)^2 \}$.

Exercise: Find the change-of-basis matrix from $\{ \mathbf{i}, \mathbf{j}, \mathbf{k} \}$ to $\{ \mathbf{j}+\mathbf{k}, \mathbf{k}+\mathbf{i}, \mathbf{i}+\mathbf{j} \}$.

Linear transformations and matrices

We have seen that $T(x) = Ax$ is a linear transformation, when A is an $m\times n$ matrix, and x is $n\times 1$. The converse is also true.

Theorem: If $T: \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, then for all v in $\mathbf{R}^n$, $T(v) = Mv$, where M is the $m\times n$ matrix whose columns are $T(e_1), T(e_2), \dots , T(e_n)$.

This theorem says that a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ "is a matrix," in the sense that it is carried out by multiplying by a matrix. Why is the theorem true? Well, each vector in $\mathbf{R}^n$ is a unique linear combination of basis vectors $\{ e_1, e_2, \dots , e_n \}$: $v = v_1e_1 + v_2e_2 + \dots + v_ne_n$. And T is linear. Thus, $T(v) = T(v_1e_1 + v_2e_2 + \dots + v_ne_n) = T(v_1e_1) + T(v_2e_2 + \dots + v_ne_n) = \dots = T(v_1e_1) + T(v_2e_2) + \dots + T(v_ne_n) = v_1T(e_1) + v_2T(e_2) + \dots + v_nT(e_n)$, using the definition of "linear transformation" over and over and over… The final equation is $T(v) = v_1T(e_1) + v_2T(e_2) + \dots + v_nT(e_n)$. Each of the $T(e_j)$ is an $m\times 1$ column vector, so the equation says that $T(v)$ is a linear combination of the columns $T(e_1)$,

$T(e_2), \ldots, T(e_n)$. That is exactly what matrix multiplication is: $(a_{ij})(v_1, v_2, \ldots, v_n) = v_1\begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \end{pmatrix}$ $+ v_2\begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{n2} \end{pmatrix} + \ldots + v_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{nn} \end{pmatrix}$. So we put the column vectors $T(e_1), T(e_2), \ldots, T(e_n)$ in a row, left-to-right, and remove all but the first and last parentheses to construct the matrix M, which we'll call $M(T)$.

Example: Find the matrix of the linear transformation of $\mathbf{R}^n$ to $\mathbf{R}^n$ that reverses the order of the entries of a vector.

Answer: $\begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \\ : & : & \ldots & : & : \\ 0 & 1 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix}$.

The spaces $\mathbf{R}^n$ are not the only vector spaces we work with. When T is a linear transformation from $\mathbf{V}$ to $\mathbf{W}$, and both $\mathbf{V}$ and $\mathbf{W}$ have finite dimension, it is useful to pick a basis for each space, and then the spaces *resemble* $\mathbf{R}^n$. We can even get a matrix representation for T, so we can say that every linear transformation between finite-dimensional spaces "is" a matrix. The purpose of this part is to show how to find that matrix, and that will it can be done! The existence of a matrix representation is true in infinite-dimensional spaces too, but bases are seldom used in such spaces. Please notice, in the argument coming up, that the finiteness of the bases in $\mathbf{V}$ and $\mathbf{W}$ is not really used! Let's let $V = \{ v_1, \ldots, v_n \}$ denote a basis of $\mathbf{V}$ and $W = \{ w_1, \ldots, w_m \}$ denote a basis of $\mathbf{W}$. So we're using n to stand for the dimension of $\mathbf{V}$, and m to stand for the dimension of $\mathbf{W}$. The first thing we'll do is to "find" the coordinates, with respect to W, of $T(v_j)$, for each j. We don't have to say how, because we know it can be done, in principle. So we get, for each j, $T(v_j) = \sum_{i=1}^{m} t_{ij}w_i$. We know that, for each v in $\mathbf{V}$, there exist unique scalars (that we've called the coefficients of v with respect to V) c_j, such that $v = c_1v_1 + \ldots + c_nv_n = \sum_{j=1}^{n} c_jv_j$. Let's substitute this into $T(v)$ in place of v. We get $T(v)$ $= T\left(\sum_{j=1}^{n} c_jv_j\right)$ = (by repeated use of the definition of linear transformation) $= \sum_{j=1}^{n} c_j \sum_{i=1}^{m} t_{ij}w_i =$ $\sum_{j=1}^{n}\sum_{i=1}^{m} c_jt_{ij}w_i = \sum_{i=1}^{m}\left\{\sum_{j=1}^{n} c_jt_{ij}\right\} w_i$. We're just about done. We recgnize that $\left\{\sum_{j=1}^{n} c_jt_{ij}\right\}$ is the i^{th} entry of $(t_{rs})C_V(v)$, and is also the i^{th} entry of $C_W(T(v))$. Therefore, to make everything into a column vector, we write $C_W(T(v)) = C_V(v)(t_{rs})$. The subscripts on the C's are there to distinguish between the two bases, V for **V**, and W for **W**. The matrix of T with respect to the pair of bases V, W is (t_{rs}). Let's denote this $M(T) = M_{V,W}(T)$ if we need to keep track of the bases. Here is a diagram that illustrates what is going on.

$$\begin{array}{ccc} \mathbf{V} & \xrightarrow{\quad T \quad} & \mathbf{W} \\ \Big\downarrow C_V & & \Big\downarrow C_W \\ \mathbf{R}^n & \xrightarrow{\quad M(T) \quad} & \mathbf{R}^m \end{array}$$

In the diagram, the arrows signify which is the domain space and the range space. You can travel thru the diagram, following the arrows, OR you can go against an arrow IF you know that the mapping for that arrow HAS AN INVERSE. In this diagram, we know that the coordinate mappings C_V and C_W have inverses, so we can do a number of trips thru the diagram. The point of it all is that if we start someplace and go to another place, we'll always get the same result, even if we follow different routes. This is expressed by saying "the diagram commutes." So, let's start with a vector v in **V**. If we go across the top, we get T(v), in **W**. Let's get from **V** to **W** another way. Start with v in V, go to $C_V(v)$ in $\mathbf{R}^n$, multiply by M(T), to get to $\mathbf{R}^m$, then apply C_W^{-1} to get to **W**. The diagram commutes, so we get to write down an equation GIVEN to be true: $C_W^{-1}(M(T)C_V(v)) =$ T(v). This is full of parentheses. Let's use the convention of writing the composition of linear mappings as multiplication. The equation then becomes $C_W^{-1}M(T)C_Vv = Tv$. Further, since 2 linear mappings that do the same thing to every vector are, by definition, equal, we can write $C_W^{-1}M(T)C_V = T$. So what? Well, we could use this equation to find T(v) in terms of coordinates. We can also us it to express M(T) in terms of T and the coordinate mappings: $M(T) = C_WTC_V^{-1}$. Making trips thru diagrams that commute is called "diagram-chasing."

Exercise: What trips thru the diagram give the last equation?

Different pairs of bases give different M(T)'s, but they *are* related, via change-of-basis matrices. Thus, if V′, W′ are other bases in **V**, **W** respectively, we know that $C_{V'} = X_{V,V'}C_V$ and $C_{W'} = X_{W,W'}C_W$. The X's are the chang-of-basis matrices. We know also that $M_{V,W}(T) = C_WTC_V^{-1}$. This equation is really the same as the one just referred to in the Exercise. Therefore, $M_{V',W'}(T) = C_{W'}TC_{V'}^{-1}$. Now, we need to substitute for the C's: $M_{V',W'}(T) = X_{W,W'}C_WTC_V^{-1}(X_{V,V'})^{-1}$. The 3 middle factors, when multiplied together, give $M_{V,W}(T)$. And $(X_{V,V'})^{-1} = X_{V',V}$ (right?). Therefore, we get the promised relation between the different matrices that represent T:

$$M_{V',W'}(T) = X_{W,W'}M_{V,W}(T)X_{V',V}.$$

Example: Let **V** denote polynomials of degree ≤ 3, and let **W** denote polynomials of degree ≤ 2. Let T denote differentiation. Thus, $T(ax^3 + bx^2 + cx + d) = 3ax^2 + 2bx + c$. Find the matrix ot T with respect to the standard bases, $V = \{x^3, x^2, x, 1\}$, of **V**, and $W = \{x^2, x, 1\}$, of **W**. Notice that this time we're putting the basis elements in

descending power order. Because we're using standard bases, C_V and C_W are easy to find: $C_V(ax^3 + bx^2 + cx + d) = (a, b, c, d)$, and $C_W(ax^2 + bx + c) = (a, b, c)$. We have a formula for $M(T)$: $M_{V,W}(T) = C_W T C_V^{-1}$. What's going on? Start with the column vector (a, b, c, d). Apply C_V^{-1}. The result is $C_V^{-1}((a, b, c, d)) = ax^3 + bx^2 + cx + d$. Apply T. The result is $3ax^2 + 2bx + c$. Apply C_W. The result is $(3a, 2b, c)$. What matrix has this output when the input is (a, b, c, d)? Here is a way to do it: We want an answer of the form $\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 3a \\ 2b \\ c \end{pmatrix}$, so we just have to find the right numbers to put in. The first row can be
$(3, 0, 0, 0)$. That will make the first entry of the output be $3a$. The second can be $(0, 2, 0, 0)$, the third $(0, 0, 1, 0)$. So a matrix that works is $\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. This is the only matrix that will work! Please think about why!

Example: Let $\mathbf{V} = \mathbf{W} = \mathbf{R}^3$. Let $V = \{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \}$, $W = \{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}$ }. Find the matrix of the identity operator with respect to V and W: $M_{V,W}(I)$. Well, it's $C_W I C_V^{-1} = C_W C_V^{-1}$. What does this say? Start with a vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ in $\mathbf{R}^3$. Then $C_V^{-1}((x, y, z)) = x\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + z\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. Then C_W transforms this to $xC_W(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}) + yC_W(\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}) + zC_W(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix})$. Now we need to find the W-coefficients of these three vectors, collect terms, and find the matrix. But this work has been done already, in a previous example! That's because of the equation $C_W = XC_V$. When we multiply on the right by C_V^{-1}, we get $C_W C_V^{-1} = X$. In that example, we found $X = \frac{1}{4}\begin{pmatrix} 6 & 10 & 6 \\ 9 & 14 & 9 \\ 11 & 16 & 9 \end{pmatrix}$. So this is the matrix of the identity operator with respect to these 2 different bases! Strange. But don't worry: if the bases are the same, the identity operator will be the familiar I.

Exercise: Show that the matrix of the identity operator on $\mathbf{V}$ is I if the 2 bases are the same.

Orientation

This is still part of the material on matrices and linear transformations, but it's important, so it gets its own heading. Which way is left, which right? Which way is up, which down? Up is not the same direction in Minneapolis as it is in the spot in the Indian Ocean, about halfway between the Kerguelen Islands and Albany, Australia, where "up" would be our "down." Our coordinate system is not the same as the standard one in Chicago, either. But they all have one thing in common: we could take a copy of our coordinate frame to Chicago, or wherever, being careful to hold it in the same direction, relative to the Earth, and then rotate it to match the one in Chicago, or wherever. But in Minneapolis or anyplace else, we still could not rotate our coordinate system to match its image in a mirror! Rotation and reflection can both be accomplished using linear transformations, but reflection of objects in ordinary space can not be accomplished. So how do we tell whether a linear transformation is “Euclidean” or not, meaning, does it represent something that can be done to a (stretchable/shrinkable) object in ordinary space, or not? The key is to look at the *determinant* of the linear transformation. So far, we only know about determinants of matrices. But if we have a linear transformation T between vector spaces $\mathbf{V}$ and $\mathbf{W}$, and bases V in $\mathbf{V}$, W in $\mathbf{W}$, we have seen that there is a matrix $M(T)$ that represents T with respect to the bases V and W. So let the spaces be the same, and let the bases be the same. We can now define “orientation-preserving.”

Definition: Let $\mathbf{V}$ be a finite-dimensional vector space with real scalars. Let T be a linear transformation that maps $\mathbf{V}$ ot $\mathbf{V}$. Let V be a basis of $\mathbf{V}$. Then T is <u>orientation-preserving</u> if the determinant of $M(T)$, the matrix of T with respect to V, V, is positive.

There is something worrisome about this definition. Maybe, if we pick another basis of V, then the matrix of T with respect to that basis might be negative! That can’t happen. So the definition is OK. Linear transformations that are orientation-preserving are the ones that can do something to a (stretchable/shrinkable) object in ordinary space. The ones with negative determinant do their thing in the mirror world. What about the ones whose determinants are 0? They squeeze the space V into a subspace of smaller dimension: they would squeeze $\mathbf{R}^3$ into a point, a line or a plane.

Note on why the definition is OK. This depends on something we haven’t studied yet: if A and B are square matrices of the same size, then $\det(AB) = \det(A)\det(B)$, where $\det(A)$ denotes the determinant of A. Since the bases are the same, $M_{V,V}(T) = C_V T C_V^{-1}$. Let W be some other basis. Then there is a square, invertible matrix X such that $C_W = XC_V$. And $M_{W,W}(T) = C_W T C_W^{-1}$. Thus, $M_{W,W}(T) = XC_V T(XC_V)^{-1} = XC_V T C_V^{-1} X^{-1} = XM_{V,V}(T)X^{-1}$. The equation we will use is: $M_{W,W}(T) = XM_{V,V}(T)X^{-1}$. This equation involves only square matrices of the same size. Let’s apply the determinant formula several times. $\det(M_{W,W}(T)) = \det(X)\det(M_{V,V}(T))\det(X^{-1}) = \det(X)\det(X^{-1})\det(M_{V,V}(T)) = \det(XX^{-1})\det(M_{V,V}(T)) = \det(I)\det(M_{V,V}(T)) = \det(M_{V,V}(T))$. We’ve also used the fact that $\det(I) = 1$. Please check it! Not ony do the determinants with respect to the different bases have the same sign, they *are* the same! Therefore, we can sensibly define the determinant of a linear transformation, even if it is not a matrix!

Definition: If $T: V \to V$ is a linear transformation on a finite-dimensional vector space, then the <u>determinant</u> of T, denoted <u>$\det(T)$</u>, is $\det(M_{V,V}(T))$, where V is any basis of V. This value is independent of the particular basis used.

Example: Find the determinant of $\frac{d}{dx} + I$, viewed as a linear transformation on the space of polynomials of degree $\leq n$, $n > 0$. We use a standard basis: $\{ x^n, x^{n-1}, x^{n-2}, \ldots, x^2, x, 1 \}$.

With respect to this basis, the matrix of $\frac{d}{dx} + I$ is $\begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ n & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 1 & 1 \end{pmatrix}$; it has 1’s on the main diagonal, and $n, n-1, n-2$, and so on, down to 1 on the diagonal below the main diagonal. It is an $(n+1)\times(n+1)$ matrix. The determinant is 1, so we can say this is the determinant of $\frac{d}{dx} + I$.

Perpendicular projection
This is still part of the material on matrices and linear transformations, but it's important, so it gets its own heading. If a particle is constrained to move in a particular line, and a force acts on the particle, only the part of that force parallel to the line can affect the particle. Here's how to find the part that's parallel to the line. First, how to make a formula that describes the line? Well, if we know the line is parallel to a vector **d**, its direction vector, and a point **s** is on the line, then any point **u** at all on the line can be expressed as $\mathbf{u} = \mathbf{s} + t\mathbf{d}$, where t is a real number. Moreover, as t varies in the real line, all the points given by this formula lie on the line. Now suppose that **F** is a vector used to portray a force. Think geometrically, or pictorially. We imagine a right triangle that has one vertex at **s**, hy-potenuse parallel to **F** and with the same magnitude as **F**, $||\mathbf{F}||$, another side parallel to **d**, with length c and direction chosen to make a right triangle. The direction is chosen so that the angle between **F** and $\pm c\mathbf{d}/||\mathbf{d}||$ is at most $\pi/2$. We choose the + sign if the angle between **F** and **d** is at most $\pi/2$, we choose – if the angle is larger than $\pi/2$ (it's never larger than π). Draw a picture, please! The length c of the side parallel to **d** has to be $||\mathbf{F}||\,|\cos\theta|$, where θ is the angle between **d** and **F**. We reccognize that $||\mathbf{F}|| \cos\theta$ is part of the formula for $\mathbf{d}\cdot\mathbf{F}$. Therefore we can write a formula for a vector **p** that will just fit that side of the triangle: $\mathbf{p} = \pm||\mathbf{F}||\,|\cos\theta|\,(\mathbf{d}/||\mathbf{d}||) = ||\mathbf{F}|| \cos\theta\,(\mathbf{d}/||\mathbf{d}||)$, because $\cos\theta$ is + (or 0) when $0 \le \cos\theta \le \pi/2$, and – when $\pi/2 < \cos\theta \le \pi$. Since $\mathbf{d}/||\mathbf{d}||$ has length 1, we recognize that $||\mathbf{F}|| \cos\theta = \mathbf{F}\cdot(\mathbf{d}/||\mathbf{d}||)$. Hence, $\mathbf{p} = (\mathbf{F}\cdot(\mathbf{d}/||\mathbf{d}||))\,(\mathbf{d}/||\mathbf{d}||)$, and this is the part of **F** that is parallel to **d**. It is a scalar multiple of **d**, and it is called the <u>orthogonal projection</u> of **F** <u>onto the subspace spanned by</u> **d**, or, for short, the <u>projection along</u> **d**. Notice that it has nothing to do with **s**. It only involves the direction of the line.

This projection is a linear transformation! It's usually written $P_{\mathbf{d}}(\mathbf{F}) = \dfrac{(\mathbf{F}\cdot\mathbf{d})\,\mathbf{d}}{||\mathbf{d}||^2}$. The other side of the right triangle is what we are thinking about when we talk about "dropping a perpendicular." It can be expressed as $\mathbf{F} - P_{\mathbf{d}}(\mathbf{F})$, or $\mathbf{F} - \dfrac{(\mathbf{F}\cdot\mathbf{d})\,\mathbf{d}}{||\mathbf{d}||^2}$. It is also a projection - the projection of **F** onto the plane perpendicular to **d**. Pictorially each can be thought of as a shadow of **F** on **d**, or on the plane perpendicular to **d**, respectively. Projections have this property: they are their own squares. Please check it out for these 2 examples! The linear transformation that sends every vector to 0, and the identity are also projections. Among numbers, the only solutions of $x^2 = x$ are 0 and 1. Among linear transformations there are many, many!

Example: Find the projection of $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ along $\mathbf{d} = 20\mathbf{i} - 4\mathbf{j} + 5\mathbf{k}$, and find the matrix of it with respect to the standard basis (used as V and as W).

We are saying, $\mathbf{p} = T(\mathbf{F})$, so $T(\mathbf{F}) = (\mathbf{F}\cdot(\mathbf{d}/||\mathbf{d}||))\,(\mathbf{d}/||\mathbf{d}||) = \dfrac{(\mathbf{F}\cdot\mathbf{d})\,\mathbf{d}}{||\mathbf{d}||^2} = (20x - 4y + 5z)\dfrac{\mathbf{d}}{||\mathbf{d}||^2} = (20x - 4y + 5z)\dfrac{20\mathbf{i} - 4\mathbf{j} + 5\mathbf{k}}{441}$. This can be expressed in matrix form directly, because we are using the standard basis. We get $C(T(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})) =$
$\dfrac{1}{441}\begin{pmatrix}20\\-5\\5\end{pmatrix}\begin{pmatrix}20 & -4 & 5\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}$, so the matrix for the transformation is $\dfrac{1}{441}\begin{pmatrix}20\\-5\\5\end{pmatrix}\begin{pmatrix}20 & -4 & 5\end{pmatrix}$. It has the form ab^t. The transformation that produces the vector perpendicular to F is $I -$
$\dfrac{1}{441}\begin{pmatrix}20\\-5\\5\end{pmatrix}\begin{pmatrix}20 & -4 & 5\end{pmatrix} = \dfrac{1}{441}\begin{pmatrix}41 & 80 & -100\\100 & -20 & -25\\-100 & 20 & -25\end{pmatrix}$. It has the form $I - ab^t$.

Example: Let T be the transformation of E^3 that rotates an arbitrary vector **w** $\pi/4$ radians counterclockwise about a fixed vector, **q**, meaning as viewed from the tip of **q** to its tail. Find the matrix of T with respect to the standard basis, used for V and W. The way we do this rotation is to imagine the plane determined by **w** and **q**, then rotate that whole plane about **q** as an axis. First, we write **q** = r**i** + s**j** + t**k**, **w** = x**i** + y**j** + z**k**. This gives us a representation of **q** and **w** in terms of **i**, **j**, and **k**. For this basis, C(**q**) = (r, s, t), and C(**w**) = (x, y, z). We don't know what the coefficients are, we just know they exist. Thinking pictorially, we notice that during this rotation, the vector **F**, that's getting turned, changes, but its projection on **q** does not change. Only its projection on the plane perpendicular to **q** changes, and what happens is that the projection on the plane perpendicular to **q** undergoes a rotation thru $\pi/4$ radians, counterclockwise, about **q**, with UP being the direction of **q**. So we need to find out how to do a rotation in a plane in $\mathbf{R}^3$. What are the vectors a**i** + b**j** + c**k** that are perpendicular to **q**? They are the ones that have dot product 0 with **q**. Thus, the plane perpendicular to **q** is the solution-set of **q**·(a**i** + b**j** + c**k**) = 0. Put in what **q** is in $\mathbf{R}^3$: (r**i** + s**j** + t**k**)·(a**i** + b**j** + c**k**) = 0 = ar + bs + ct = 0. The vectors (b, –a, 0) and (0, –c, b) form a basis if all of a, b, and c are non-zero. Let's assume that for the moment. Please check that those vectors do form a basis! What we do is this: find a vector that is perpendicular to each of (a, b, c) and (b, –a, 0), and so that the three vectors follow the right-hand rule. This is what the cross product does, so we'll use that: $(a, b, c)\times(b, -a, 0) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & b & c \\ b & -a & 0 \end{vmatrix} = \mathbf{i}ca - \mathbf{j}(-bc) - \mathbf{k}(a^2 + b^2)$ "=" $(ac, bc, -(a^2 + b^2))$ in $\mathbf{R}^3$. Although E^3 and $\mathbf{R}^3$ are technically different, they are usually freely interchanged, as here! Now we have a nice basis - all that remains is to divide each vector by its length, and then we'll have a nicer basis: it consists of mutually perpendicular unit vectors arranged in the same order that **i**, **j**, and **k** are ($\begin{vmatrix} a & b & c \\ b & -a & 0 \\ ac & bc & -(a^2+b^2) \end{vmatrix} > 0$; check it, please!). It it a basis that is set up to do what we want: rotate in the plane of the second 2 vectors, counterclockwise from the first vector's point of view. We can do this rotation with a rotation matrix: $\begin{pmatrix} \cos\pi/4 & -\sin\pi/4 \\ \sin\pi/4 & \cos\pi/4 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. We work on a coordinate vector (u, v, w), acting only on the second 2 coordinates (we can do this using block matrix ideas): the rotated vector is $\left(u, \frac{v-w}{\sqrt{2}}, \frac{v+w}{\sqrt{2}}\right)$. Now we take these coordinates, multiply times the 3 vectors we had, divided by their lengths, arrange so u, v, and w are separated, to get the answer. And the answer is,

$$T\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{vmatrix} a/A & (b/B)+(ac/C) & (-b/B)+(ac/C) \\ b/A & (-a/B)+(ac/C) & (a/B)+(bc/C) \\ c/A & -(a^2+b^2)/C & -(a^2+b^2)/C \end{vmatrix},$$

where $A = \sqrt{a^2+b^2+c^2}$, $B = \sqrt{2a^2+2b^2}$, and $C = \sqrt{2(a^2+b^2+c^2)(a^2+b^2)}$. At least, this is the answer in the new basis! But we want the answer in terms of the original basis! See if you can do it!

If $Ax = \lambda x$, and $x \neq 0$, then λ is called an <u>eigenvalue</u> of A, and x an <u>eigenvector (belonging to λ)</u>. We'd like to have a basis of $\mathbf{R}^n$ consisting of eigenvectors of A, because then the effect of A on a vector x, expressed in terms of the basis of eigenvectors, would just be to multiply the i^{th} coordinate of x, with respect to the basis of eigenvectors, by the i^{th} eigenvalue. **Example**: $A = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}$. Let $p = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $Ap = \begin{pmatrix} 4 \\ 4 \end{pmatrix} = 4\begin{pmatrix} 1 \\ 1 \end{pmatrix} =$ 4p, so p is an eigenvector of A that belongs to the eigenvalue 4. Let $q = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$. Then $Aq = \begin{pmatrix} 1 \\ -2 \end{pmatrix} = q = 1q$, so q is an eigenvector of A that belongs to the eigenvalue 1. Since p and q are not proportional, {p, q} is a linearly independent set in $\mathbf{R}^2$, so it is a basis. Let's make a matrix out of p and q: $S = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Then $AS = \begin{pmatrix} 4 & 1 \\ 4 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} =$ SD, where
$D = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$. The columns of S comprise a linearly independent set, so S is invertible; $S^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Multiply the equation AS

The general quadratic equations in 2 variables, x, y, whose solution sets are known as conic sections, can be classified using the tools of linear algebra. The main steps are:

(1) Write the quadratic equation in matrix form; use a symmetric matrix.

(2) Find the eigenvalues and eigenvectors of the symmetric matrix. If all you want to know is what kind of conic section you have, just find the eigenvalues. If you need to know whether or not it's degenerate, keep going.

(3) Make a diagonal matrix, D, that has the eigenvalues as its diagonal entries: $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

(4) Make a square matrix, S, out of your eigenvectors by first putting each one into the column corresponding to the eigenvalue it belongs to, then divide each column by its length. This works if the eigenvalues are different. If they're not, there is another step. It will be listed at the end, as step (4b), in case it's needed. As a check, the product SS^t should be I.

(5) Define new coordinates, x', y' in terms of x, y by $\begin{pmatrix} x' \\ y' \end{pmatrix} = S^t\begin{pmatrix} x \\ y \end{pmatrix}$. Note that this means, for the special S's whose transposes are their inverses, that $\begin{pmatrix} x \\ y \end{pmatrix} = S\begin{pmatrix} x' \\ y' \end{pmatrix}$.

(6) Substitute the new coordinates into the vector form of the equation. There is a shortcut, that will come to light in the example.

(7) Complete any squares that can be completed. This can only be done if the corresponding eigenvalue is not zero. This will give you new coordinates that involve a translation. Call them x'' and y''. Your quadratic equation will look like $\lambda_1(x'')^2 + \lambda_2(y'')^2 = K$ if neither eigenvalue is 0, like $\lambda_1(x'')^2 + k(y'') = K$ if $\lambda_2 = 0$, and like $kx'' + \lambda_2(y'')^2 + = K$ if $\lambda_1 = 0$, where k and K are constants whose values depend on the preceding steps. Now you can use the standard techniques for classifying your quadratic. The case $\lambda_1 = 0 = \lambda_2$ can't arise unless your quadratic has NO second-degree terms. In that case, I wouldn't call it a quadratic, but it has to be included for completeness, unless you want to specify that
at least one of A, B, and C is non-zero.

(8) The last step, if needed, is to sketch the solution-set in the xy-variables. This can be done by substituting back, for x' and y' in terms of x and y, using the first formula in (5).

(4b) If $\lambda_1 = \lambda_2$, your eigenvectors might not be perpendicular to each other. You can choose one of them as the first one, call it v_1. Then replace v_2 by $v_2 - \frac{v_2 \cdot v_1}{\| v_1 \|} v_1$, and call this new vector v_2 too. Now $v_1 \cdot v_2 = 0$. This is part of the Gram-Schmidt Orthogonalization Process, and it works no matter what size or even what kind of vectors you have. But with 2-vectors, all you have to do is switch the coordinates and multiply one of them by -1 to get a perpendicular vector!

Example: $10x^2 + 24xy + 3y^2 + x - 12y + 15 = 0$.

(1) Write in matrix form: $(x\ y)\begin{pmatrix} 10 & 12 \\ 12 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (1\ -12)\begin{pmatrix} x \\ y \end{pmatrix} + 15 = 0$.

(2) Find eigenvalues and eigenvectors - that is, find λ such that $\begin{pmatrix} 10-\lambda & 12 \\ 12 & 3-\lambda \end{pmatrix}$ has a non-trivial kernel: choose $x = -12$ and $y = 10-\lambda$; this choice makes the first entry of $\begin{pmatrix} 10-\lambda & 12 \\ 12 & 3-\lambda \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0$, so the problem is to pick λ so that the second entry, $12x + (3-\lambda)y$, is zero. Now, $12x + (3-\lambda)y = -144 + (3-\lambda)(10-\lambda) = -144 + 30 - 13\lambda + \lambda^2 = \lambda^2 - 13\lambda - 114 = (\lambda - 19)(\lambda + 6) = 0$ means the eigenvalues are 19 and -6. So we get 2 eigenvectors. One that belongs to 19 has $x = -12$, $y = 10-\lambda = 10-19 = -9$. The other eigenvector MUST be perpendicular to this, and the trick in (4b) suggests that an eigenvector for -6 is $x = -9$, $y = 12$. Please multiply each of these by our matrix to see that each one actually is an eigenvector! Now, we can divide by -3, so let's use $(4, 3)$ as 19's eigenvector, and $(3, -4)$ as -6's eigenvector.

(3) Make a diagonal matrix, D, that has the eigenvalues as its diagonal entries: $D = \begin{pmatrix} -6 & 0 \\ 0 & 19 \end{pmatrix}$. Note that we're choosing -6 as the first eigenvalue, 19 as the second one.

(4) Make a square matrix, S, out of your eigenvectors by first putting each one into the column corresponding to the eigenvalue it belongs to, then divide each column by its length: first version of
$S = \begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}$. The length of each column is 5, so just divide the whole matrix by 5, to get the final version of $S = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}$. Please check now that $S^{-1} = S^t$.

(5) Define new coordinates, x', y' in terms of x, y by $\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$.

(6) Substitute the new coordinates into the vector form of the equation: use $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}$.

$$(x\ y)\begin{pmatrix} 10 & 12 \\ 12 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (1\ -12)\begin{pmatrix} x \\ y \end{pmatrix} + 15 =$$
$$= \left(\frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}\right)^t \begin{pmatrix} 10 & 12 \\ 12 & 3 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + (1\ -12)\frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + 15$$
$$= \frac{1}{25}(x'\ y')\begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} 10 & 12 \\ 12 & 3 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + (1\ -12)\frac{1}{5}\begin{pmatrix} 3 & 4 \\ -4 & 3 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + 15$$

$$= \frac{1}{25}(x'\ y')\begin{pmatrix}-150 & 0\\ 0 & 475\end{pmatrix}\begin{pmatrix}x'\\ y'\end{pmatrix} + \frac{1}{5}(49\ -32)\begin{pmatrix}x'\\ y'\end{pmatrix} + 15$$
$$= (x'\ y')\begin{pmatrix}-6 & 0\\ 0 & 19\end{pmatrix}\begin{pmatrix}x'\\ y'\end{pmatrix} + \frac{49}{5}x' - \frac{32}{5}y' + 15$$
$$= -6(x')^2 + 19(y')^2 + \frac{49}{5}x' - \frac{32}{5}y' + 15 = 0.$$

The shortcut mentioned earlier is this: you don't have to do the triple matrix product! You can just write down the sum of the eigenvalues times the squares of the corresponding new coordinates. The only thing to be careful about is to make sure you match them properly! Now we can tell that this looks like a hyperbola.

(7) Complete any squares that can be completed. Both squares can be completed. I get, at first

$$-6\left(x' - \frac{49}{60}\right)^2 + 19\left(y' - \frac{32}{190}\right)^2 + 15 + 6\left(\frac{49}{60}\right)^2 - 19\left(\frac{32}{190}\right)^2 = 0,$$

then

$$-6\left(x' - \frac{49}{60}\right)^2 + 19\left(y' - \frac{32}{190}\right)^2 + \frac{8419}{456} = 0,$$

and, finally, with $x'' = x' - \frac{49}{60}$, and $y'' = y' - \frac{32}{190}$,

$$-6(x'')^2 + 19(y'')^2 + \frac{8419}{456} = 0.$$

From this we can tell that, *in* $x'y'$*-space*, when we put this equation in standard form, we will get a hyperbola that opens to the left and right, as $(x'')^2 - (y'')^2 = 1$ does.

(8) Write the equation in xy-space, using the formulas for x' and y' in terms of x and y. This step is left to you to do. It amounts to a rearrangement of the original equation, into a form that reveals the nature of the solution-set.

1. The **dot product** of two vectors, a, b, in $\mathbf{R}^n$ is denoted a•b, and is defined by $a \cdot b = \sum_{i=1}^{n} a_i b_i$. This is an example of an operation that has 2 vectors as inputs, and yields a number as output. It is also known as the inner product, and as the scalar product, of the vectors a and b.

(a) Verify that, in $\mathbf{R}^n$, the dot product has these properties:

(1) For all a in $\mathbf{R}^n$, $a \cdot a \geq 0$, and, if $a \cdot a = 0$, then $a = 0$ (the vector);

(2) For all a, b in $\mathbf{R}^n$, $a \cdot b = b \cdot a$ (in cases when we use complex scalars, e.g. in EE, this gets changed to: $a \cdot b = \overline{b \cdot a}$);

(3) For all a, b, c in $\mathbf{R}^n$, $(a+b) \cdot c = a \cdot c + b \cdot c$;

(4) For all a, b in $\mathbf{R}^n$, and for all scalars α, $(\alpha a) \cdot b = \alpha(a \cdot b)$.

We define the length of a to be $\sqrt{a \cdot a}$. We denote it $|a|$, or sometimes $||a||$. Especially in the latter case, we call it by the (pretentious?) name norm of a. So, (1) says that the length of a vector is non-negative, and that a vector with zero length is the zero vector; (2) says the dot product is symmetric (conjugate symmetric, in the complex-scalar case); (3) and (4) together say that the dot product is linear in the first variable (when the second variable is held fixed); (3) says it is additive in the first variable, (4) says it is homogeneous (of degree one) in the first variable.

(b) Verify, using only the properties (1) - (4) in (a), that, in $\mathbf{R}^n$, the dot product has these properties:

(1′) It is linear in the second variable (when the first variable is held fixed);

(2′) Length is positively homogeneous (of degree one): For all a in $\mathbf{R}^n$, and for all scalars α, $||\alpha a|| = |\alpha|\ ||a||$;

(3′) The Parallelogram Identity is true: For all a, b in $\mathbf{R}^n$, $||a + b||^2 + ||a - b||^2 = 2||a||^2 + 2||b||^2$;

(4′) The zero vector has length 0;

(5′) The Schwartz Inequality is true: For all a, b in $\mathbf{R}^n$, $|a \bullet b| \le ||a||\ ||b||$;

(6′) The triangle inequality is true: For all a, b, c in $\mathbf{R}^n$, $||a - c|| \le ||a - b|| + ||b - c||$.
Actually, you may need to use some of the basic vector-space properties, too, and that's OK. The only one of these that I expect you'll find very hard to do is the Schwartz Inequality, (5′). Here are 2 possible approaches: You could minimize $g(t) = ||a - tb||^2$; what do we know about the discriminant in a quadratic that is never negative? You could show that the Schwartz Inequality is true when one of a and b is 0, then true when both a and b have length 1, then use the fact that, if a vector a is not the zero vector, then $\frac{a}{||a||}$ is a vector that has length 1, and then use positive homogeneity.

(c) In $\mathbf{R}^n$, we DEFINE the angle θ between two vectors *in terms of its cosine*, by the equation $a \bullet b = ||a||\ ||b|| \cos \theta$. We always regard this angle as being between 0 and π, inclusive. We say that two vectors a and b are perpendicular, or (more pretentiously?) orthogonal, if $a \bullet b = 0$. Thus, the angle between orthogonal vectors is $\pi/2$.

In $\mathbf{R}^n$, a hyperplane is defined to be the solution-set, H, of an equation of the form $a \bullet (x - p) = 0$, where $||a|| > 0$. That is, a hyperplane is the set of all vectors that make the equation true: $H = \{ x : a \bullet (x - p) = 0 \}$. Notice that $p \in H$.

(c-1) Show that, if x and y are in the hyperplane $a \bullet (x - p) = 0$, that $x - y$ is orthogonal to a. The vector a is said to be perpendicular to the hyperplane; it is called a normal (vector) to H.

(c-2) Find an equation for the (hyper)plane H_2 in $\mathbf{R}^3$ that passes thru the three points (1, 0, 0), (0, 1, 0), and (0, 0, 1); this means to find a and p such that $H_2 = \{ x : a \bullet (x - p) = 0 \}$.

(c-3) Find an equation for the (hyper)plane H_3 in $\mathbf{R}^3$ that passes thru the three points (2, –1, 0), (–1, 2, –1), and
(0, –1, 2);

(c-4) Let $a = (1, 2, 3)$, $b = (3, -1, 2)$. Set up a system of equations, and solve it, to find all vectors in $\mathbf{R}^3$ that are perpendicular to both a and b.

(c-5) A parametric equation for a line in $\mathbf{R}^n$ is $P + tD$, where P, D are vectors in $\mathbf{R}^n$, and $D \ne 0$. Find a parametric equation for the line in $\mathbf{R}^3$ that is perpendicular to the plane H_3 in (c-3), and passes thru the origin.

(c-6) Suppose that L_1 and L_2 are two different lines in $\mathbf{R}^3$. Is it always possible to find a (hyper)plane P that contains both lines? If so, say how to find one; if not, find the condition the lines must satisfy in order for the answer to be YES. Apply your work to this pair of lines: the ones in (c-4) and (c-5).

2. This is about ordinary differential equations with constant coefficients, and systems of ordinary differential equations with constant coefficients. Examples: $y'' + y = 0$, $y'' + y = 1$, $y'' + y = t$, $y'' + y = e^t$, $y'' + y = \sin t$, $y'' + y = t \sin t$. We are going to use t as the independent variable instead of x, to connote *time*, but you can use x if you wish. We "have to" use complex numbers as scalars in this application! We *could* get around the use of complex numbers, but it would make things harder to DO. **Here is the idea**: try a solution of the form $y = A\,t^n e^{zt}$, where A is a complex number, n is a non-negative integer, and z is a complex number - it could be 0, so this means we can try polynomials as solutions. At first, try $y = e^{zt}$ first. Thus, in $y'' + y = 0$, plugging in $y = e^{zt}$ leads to $(z^2 + 1)e^{zt} = 0$. This means $= 0$ *for all* t. This can only happen if $z^2 + 1 = 0$. Thus, $z = \pm i$. This gives 2 solutions, e^{it} and e^{-it}. We'd like real solutions. Since a linear combination of solutions is a solution, we get $\cos t = \frac{1}{2}(e^{it} + e^{-it})$, and $\sin t = \frac{1}{2i}(e^{it} - e^{-it})$, so $\cos t$ and $\sin t$ are solutions. Now look at $y'' + y = 1$. To find the general solution, we will look for a particular solution and add the general homogeneous solution, $a\cos t + b\sin t$, to it. Try $y = e^{zt}$. We get that $(z^2 + 1)e^{zt} = 1$. This is supposed to be true for all t. This can only happen if $z = 0$. Thus, a particular solution is $y(t) = 1$ for all t. The general solution is $y = 1 + a\cos t + b\sin t$. For $y'' + y = t$, try $y = A\,te^{zt}$. We get $A\,te^{zt}(z^2 + 1) + 2zA\,e^{zt}$; for this to equal t for all t, we need $z = 0$, $A = 1$. Thus, a particular solution is $y(t) = t$ for all t, so the general solution is $t + a\cos t + b\sin t$. So it seems that, if the RHS has a term $t^n e^{zt}$, it's a good idea to try $y = A\,t^n e^{zt}$. For $y'' + y = e^t$, try $y = A\,e^t$. We will get $A(1^2 + 1)e^t$; so choose $A = 1/2$ to get a particular solution. The general solution is $y = \frac{1}{2}\,e^t + a\cos t + b\sin t$. But *if the RHS has a term in it that is a solution of the homogeneous equation, we need to multiply by one or more extra powers of* t. Thus, for $y'' + y = \sin t$, try $A\,t\sin t$. You will find that it doesn't work. Try $A\,t\cos t$. You get $-2A\sin t$, so $A = -\frac{1}{2}\cos t$ gives a particular solution. The general solution is $-\frac{1}{2}\cos t + a\cos t + b\sin t$. It's easier, in the long run, to try for a particular solution in the form $Q(t)\,e^{zt}$, where Q is a polynomial in t of the right degree, coefficients undetermined, even when z is a complex number. For $y'' + y = t\sin t$, try $y = (At^2 + Bt + C)e^{it}$. Think of the LHS of the differential equation as a linear transformation of the form $P(D) = (D^2 + 1)$, acting on y, where D is short for $\frac{d}{dt}$. So we seek a solution of the equation $P(D)y = te^{it}$, and you plug in $y = (At^2 + Bt + C)e^{it}$ You get $P(D)(At^2 + Bt + C)e^{it}$, and it's a bore to do all the differentiations! So you notice that $P(D)e^{zt} = P(z)e^{zt}$! And you try differentiating this equation WITH RESPECT TO z, treating t as a constant, just as you treated z as a constant when you did all those differentiations with respect to t. So $P(D)$ is a constant as far as z is concerned, so the equation reads $P(D)te^{zt} = P'(z)e^{zt} + P(z)te^{zt}$. In our example, $z = i$, and $P(z) = (z^2 + 1)$, so $P(i) = 0$, and we get $P(D)te^{it} = P'(i)e^{it} = 2ie^{it}$. Notice that we can't plug in the i until AFTER we differentiate with respect to z. Notice that $P(D)Ate^{it} = 2iAe^{it}$ can easily be split into real and imaginary parts if we choose A to be real. When that is done, we get $P(D)At\cos t = -2A\sin t$, and $P(D)At\sin t = 2A\cos t$. This is relevant for an earlier example. Now, we can differentiate more than once with respect to z, and we get $P(D)t^2e^{zt} = P''(z)e^{zt} + 2P'(z)te^{zt} + P(z)t^2e^{zt}$. For the present example this reads $P(D)t^2e^{it} = 2e^{it} + 4ite^{it} + 0\,t^2e^{zt}$. There is an extra "lower-order" term on the right, but we can add ite^{it} and get rid of it: $P(D)(t^2e^{it} + ite^{it}) = 4ite^{it}$. Now we can divide both sides by $4i$, and take real and imaginary parts of the solution, to find a particular solution for the last example.

(a) Find the general solution of $y''' + y = t\cos(2\pi t/3)$;

(b) Find the solution of $y''' + 2y'' - 2y' + 3y = 0$ such that $y(0) = 3$, $y'(0) = 1$, $y''(0) = -2$;
(c) Under what conditions on the coefficients on the LHS in $my'' - cy' - mgy = \cos \pi t$, with $y(0) = 1$, $y'(0) = -1$, will the solution $y(t)$ remain bounded as $t \to \infty$? Here, m, c, and g are positive constants.
(d) Find the vector-valued function $y(t) = (y_1(t), y_2(t), y_3(t))$ that satisfies the <u>system of differential equations</u>

$$\begin{aligned} 2y_1' - y_2 &= 0, \\ -y_1 + 2y_2' - y_3 &= 1, \\ -y_2 + 2y_3' &= t, \end{aligned}$$

and such that $y(0) = (0, 1, 2)$.
Suggestion: Write Dy_i for y_i' and do Gauss elimination, treating D like an unknown number, but use "fraction-avoiding" methods, since we can't divide by D. The reduced matrix will then have entries that can be interpreted as differential operators, and you backsolve by solving a bunch of *single* differential equations.

3. An often encountered kind of linear transformation is one that maps a vector space V into itself. In particular, if $T(x) = Ax$ is defined by multiplying the vector x by the matrix A, then A must be a square matrix. In $\mathbf{R}^n$, what does a matrix "do" to a vector? If the matrix is a multiple of I, then it simply stretches x, or shrinks it, perhaps reversing its direction at the same time. To include all these possibilities in one term we say $A = cI$ is a <u>dilation</u>. If the matrix A is <u>diagonal</u>, meaning that all its off-diagonal entries are 0, then Ax multiplies each coordinate of x by a possibly different number. If the columns of A all have length 1, and if they are perpendicular to each other, the matrix A is called an <u>orthogonal matrix</u>; A then rotates, perhaps reflects, x.
(a) Show that, if a matrix A is orthogonal, then $\|Ax\| = \|x\|$, and also, $Ax \cdot Ay = x \cdot y$, for all x, y in $\mathbf{R}^n$. That is, an orthogonal matrix <u>preserves angles</u> (between pairs of vectors);
(b) Show that, if a matrix A is orthogonal, then A^t is the inverse of A, and vice versa;

(c) Sometimes a matrix acts like a dilation on some, but not all vectors. If $Ax = \lambda x$, and $x \neq 0$, then λ is called an <u>eigenvalue</u> of A, and x an <u>eigenvector (belonging to λ)</u>. We'd like to have a basis of $\mathbf{R}^n$ consisting of eigenvectors of A, because then the effect of A on a vector x, expressed in terms of the basis of eigenvectors, would just be to multiply the i^{th} coordinate of x, with respect to the basis of eigenvectors, by the i^{th} eigenvalue.

Example: $A = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}$. Let $p = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $Ap = \begin{pmatrix} 4 \\ 4 \end{pmatrix} = 4\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 4p$, so p is an eigenvector of A that belongs to the eigenvalue 4. Let $q = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$. Then $Aq = \begin{pmatrix} 1 \\ -2 \end{pmatrix} = q = 1q$, so q is an eigenvector of A that belongs to the eigenvalue 1. Since p and q are not proportional, $\{p, q\}$ is a linearly independent set in $\mathbf{R}^2$, so it is a basis. Let's make a matrix out of p and q: $S = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Then $AS = \begin{pmatrix} 4 & 1 \\ 4 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} = SD$, where $D = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$. The columns of S comprise a linearly independent set, so S is invertible; $S^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Multiply the equation $AS = SD$ on the right by S^{-1}; the result is $A = SDS^{-1}$. So what? Well, D^{-1} is easy to calculate! So form $SD^{-1}S^{-1} = \frac{1}{4}\begin{pmatrix} 2 & -1 \\ -2 & 3 \end{pmatrix}$ and multiply times A: $ASD^{-1}S^{-1} =$
$SDS^{-1}SD^{-1}S^{-1} = SDD^{-1}S^{-1} = SS^{-1} = I$, so we can find the inverse easily. Now take the

square root of D: it's $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. Now form $S\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}S^{-1} = \frac{1}{3}\begin{pmatrix} 5 & 1 \\ 2 & 4 \end{pmatrix}$, and square it (you do it the usual way, please!):

$S\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}S^{-1}S\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}S^{-1} = S\begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}S^{-1} = A$. We have found a square root of A. We can calculate $e^{tA} = \sum_{n=0}^{\infty}\frac{(tA)^n}{n!}$, not by summing a series of matrices, but in the same way we found a square root: Form e^{tD}, and put the result between S and S^{-1}: $S\begin{pmatrix} e^{4t} & 0 \\ 0 & e^{t} \end{pmatrix}S^{-1} =$

$\frac{1}{3}\begin{pmatrix} 2e^{4t}+e^{t} & e^{4t}-e^{t} \\ 2e^{4t}-2e^{t} & e^{4t}+2e^{t} \end{pmatrix}$. It turns out that $y(t) = e^{tA}\begin{pmatrix} u \\ v \end{pmatrix}$ is the solution of $\frac{dy}{dt} = Ay$ having $y(0) = \begin{pmatrix} u \\ v \end{pmatrix}$. So IF we can find the eigenvalues and thier eigenvectors of a matrix A, we can express A in a form that makes many tasks easier. Unfortunately, it can't always be done. But it CAN be done (in principle) if there are n distinct eigenvalues, where $n\times n$ is the size of A.

(c-1) Let $A = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$. Find the eigenvalues and eigenvectors of A, find A^{-1}, *four* solutions, X, of $X^2 = A$, and the solution of $\frac{dy}{dt} = Ay$ having $y(0) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

(c-2) Let $A = \frac{1}{18}\begin{pmatrix} 62 & -16 & -4 \\ 17 & 29 & -13 \\ 29 & -13 & 17 \end{pmatrix}$. Find the eigenvalues and eigenvectors of A, and the solution of $\frac{dy}{dt} = Ay$ having $y(0) = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

(c-3) Let $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$. Find the eigenvalues and eigenvectors of A, find A^{-1}, *eight* solutions, X, of $X^2 = A$, and

the solution of $\frac{dy}{dt} = Ay$ having $y(0) = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$.

discussion of range - image, null-space - kernel
examples that are matrices, then examples that are not, esp ODEs;
identifying spaces depends on how they are defined
determinants I: how-to DONE
discussion of vectors in the plane and in space,
change of basis, frames of reference - matrices, space, ODEs, functions
bilinear forms, quadratic forms, quadratic equations and conic sections
symmetric matrices, eigenvalues, eigenvectors;
inner products; applied to matrices, to space and to functions:
relation between linear transformations and matrices, on finite-dimensional spaces
linear trans determined by what it does to a basis
constructing linear transformations that "do" various things
permutation matrices
determinants II: what they are
uselessness of basis in infinite dimensional case
direct sum of vector spaces, graph of a lin trans is a subspace of direct sum
ease of changing bases with rotations, examples in 3-space

polar coordinates in 2-space and 3-space; n-space?
self-adjoint ODE problems: Bdry-value problems
row rank=column rank
rank+nullity=dimension of domain
orthogonal complements, orthogonal projections
Fourier series, summing $1/n^2$, orthogonal projections,
every matrix is a sum of rank-one matrices
Finding eigenvectors for operators of the form I + small rank
Finding inverses of operators of the form I + small rank
looking at Gauss elimination in terms of operators
Finding a maximal linearly independent subset of a set of column vectors - Sam's approach, standard approach

So here is what I'll do: Make a basis for E^3 that looks like **i**, **j**, and **k**, namely, the vectors in the basis have unit length and are mutually perpendicular, and the first 2 vectors in this basis will form a basis for the plane perpendicular to **q**, while the third will be a multiple of **q**. I'll call them $\mathbf{i}'$, $\mathbf{j}'$, and $\mathbf{k}'$. So, I'll let $\mathbf{i}' = \frac{1}{\sqrt{a^2+b^2}}\begin{pmatrix} b \\ -a \\ 0 \end{pmatrix}$. I'd like to do the same for $\begin{pmatrix} 0 \\ -c \\ b \end{pmatrix}$, but it's not perpendicular to $\mathbf{i}'$. So I'll subtract from it its projection on $\mathbf{i}'$, and that will be perpendicular to $\mathbf{i}'$. The formula is $P_{\mathbf{i}'}(\mathbf{F}) = \frac{(\mathbf{F}\cdot\mathbf{i}')\,\mathbf{i}'}{||\mathbf{i}'||^2}$, for the projection on $\mathbf{i}'$. Since $\mathbf{i}'$ has length 1, this simplifies to $(\mathbf{F}\cdot\mathbf{i}')\,\mathbf{i}'$. Use $\begin{pmatrix} 0 \\ -c \\ b \end{pmatrix}$ for **F**, and subtract the result from $\begin{pmatrix} 0 \\ -c \\ b \end{pmatrix}$: $\begin{pmatrix} -\frac{abc}{a^2+b^2} \\ -\frac{b^2c}{a^2+b^2} \\ b \end{pmatrix}$ is what I get.