NOTES ON LINEAR ALGEBRA

CONTENTS:

[1] MULTIPLYING MATRICES
[2] GAUSSIAN ELIMINATION
[3] INVERTING MATRICES

[1] MULTIPLYING MATRICES:

For ease of presentation, I will NOT draw the parentheses around the matrices correctly. If I were to, I'd have to use either the Equation Editor (which takes more time) or LaTeX (which your computer has trouble reading).

Let's say we have the matrix A =

(1 2)
(3 4)

And we want to multiply it by the column vector v =

(5)
(6)

The answer is Av =

(1 2) (5) (1*5 + 2*6) (17)
(3 4) (6) = (3*5 + 4*6) = (39)

Let's do another example. Let B =

(2 7)
(3 5)

and let the vector w =

(1)
(3)

Then B w =

$$\begin{pmatrix} 2 & 7 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2*1 + 7*3 \\ 3*1 + 5*3 \end{pmatrix} = \begin{pmatrix} 23 \\ 18 \end{pmatrix}$$

Let's study a bigger matrix now. Let C =

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

and consider the vector x =

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Then C x =

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1*2 + 2*1 \\ 3*2 + 4*1 \\ 5*2 + 6*1 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \\ 16 \end{pmatrix}$$

And finally, let's look at D =

$$\begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 2 \\ 5 & 6 & 3 \end{pmatrix}$$

and the vector y =

$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

Then the product D y =

$$\begin{pmatrix} \mathbf{1 & 2 & 0} \\ \mathbf{3 & 4 & 2} \\ \mathbf{5 & 6 & 3} \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{2} \end{pmatrix} = \begin{pmatrix} \mathbf{1*1 + 2*0 + 0*2} \\ \mathbf{3*1 + 4*0 + 2*0} \\ \mathbf{5*1 + 6*0 + 3*2} \end{pmatrix}$$

This is basically how to multiply a matrix by a column vector. Now we want to study how to multiply two matrices together. We have the following rule, which we proved:

Matrix Multiplication Rule:

Let's say A has R_a rows and C_a columns, and B has R_b rows and C_b columns. This means A is an R_a x C_a matrix, and B is an R_b x C_b matrix. Then we can do the multiplication AB if and only if $C_a = R_b$, and the resulting matrix AB has R_a rows and C_b columns.

For example, if A is 3x4 and B is 4x2, then we can do the multiplication AB, and the product AB is a 3x2 matrix; however, we cannot do the multiplication BA, for 2 ≠ 3.

Let's do some examples: Let the matrices A and B be (respectively)

(1 2)	and	(5 6)
(3 4)		(7 8)

Then in this case we can multiply in EITHER order, as both are 2x2. Let's do AB =

(1 2) (5 6)
(3 4) (7 8)

The way we multiply matrices is column by column. To find the first column in the product, we multiply the matrix A by the first column of B, and that's the answer. To find the second column of the product, we multiply A by the second column of B.

Step 1: Finding the first column of the product:

(1 2) (5) (1*5 + 2*7) (19)
(3 4) (7) = (3*5 + 4*7) = (43)

Step 2: Finding the second column of the product:

(1 2) (6) (1*6 + 2*8) (22)
(3 4) (8) = (3*6 + 4*8) = (50)

Step 3: Combining the above:

$$\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}5 & 6\\ 7 & 8\end{pmatrix} = \begin{pmatrix}19 & 22\\ 43 & 50\end{pmatrix}$$

Let's do a harder one: Let the matrices C and D be (respectively)

$$\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 2 & 1 & 0\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}3 & 0\\ 1 & 2\\ 0 & 5\end{pmatrix}$$

First, let's check and make sure we can multiply CD. C is 3x3, D is 3x2, so yes we can, and the product will be 3x2.

Step 1: C times the first column of D gives the first column of CD

$$\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 2 & 1 & 0\end{pmatrix}\begin{pmatrix}3\\ 1\\ 0\end{pmatrix} = \begin{pmatrix}1*3 + 2*1 + 3*0\\ 4*3 + 5*1 + 6*0\\ 2*3 + 1*1 + 0*0\end{pmatrix} = \begin{pmatrix}5\\ 17\\ 7\end{pmatrix}$$

Step 2: C times the second column of D gives the second column of CD

$$\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 2 & 1 & 0\end{pmatrix}\begin{pmatrix}0\\ 2\\ 5\end{pmatrix} = \begin{pmatrix}1*0 + 2*2 + 3*5\\ 4*0 + 5*2 + 6*5\\ 2*0 + 1*2 + 0*5\end{pmatrix} = \begin{pmatrix}19\\ 40\\ 2\end{pmatrix}$$

Step 3: Combining the above yields CD =

$$\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 2 & 1 & 0\end{pmatrix}\begin{pmatrix}3 & 0\\ 1 & 2\\ 0 & 5\end{pmatrix} = \begin{pmatrix}5 & 19\\ 17 & 40\\ 7 & 2\end{pmatrix}$$

[2] GAUSSIAN ELIMINATION:

Matrices can be used to represent systems of equations, which we then try to solve. For example, let's say we have the two equations:

$$3x + 2y = 5$$
$$4x + 5y = 7$$

Then we can write this in matrix form by

$$\begin{pmatrix} 3 & 2 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

Or, if we had the three equations

$$3x + 2y + 5z = 8$$
$$2x + 2y + 4z = 7$$
$$7x + 9y + 0z = 1$$

Then we can write this in matrix form by

$$\begin{pmatrix} 3 & 2 & 5 \\ 2 & 2 & 4 \\ 7 & 9 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 8 \\ 7 \\ 1 \end{pmatrix}$$

Now, we want to find a way to solve such systems of equations. Let's start with an easy example:

$$1x + 2y = 1$$
$$3x + 7y = 2$$

We can write this in matrix form by

$$\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Now, let's look at the two equations. If we multiply the first equation by -3 we get: $-3x - 6y = -3$. If we then add this to the second equation ($3x + 7y = 2$) we get a new second equation:

$$
\begin{array}{rr}
 & 3x + 7y = \ \ 2 \\
+ & -3x \ - 6y \ = -3 \\
 & \text{----------} \\
 & 0x \ + 1y = -1
\end{array}
$$

So now we have the two equations

$$
\begin{array}{l}
1x + 2y = \ 1 \\
0x + 1y = -1
\end{array}
$$

which we can write in matrix form as

$$
\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
$$

We started with the matrix

$$
\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}
$$

If we multiply the first row by -3 and add that to the second row, we get the matrix

$$
\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}
$$

And if we multiply 1 by -3 and add it to 2 we get the vector

$$
\begin{pmatrix} 1 \\ -1 \end{pmatrix}
$$

So we see we can symbolically represent multiplying and adding equations by multiplying and adding rows. Slowly, here goes:

The goal is to reduce the matrix to something easy to work with, namely something with all zeros below the diagonal.

We start with

$$\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (1) \\ (2) \end{pmatrix}$$

Step 1: What do we need to multiply the first row by to cancel the 3 in the second row? Or, find 'a' such that 1a + 3 = 0, hence a = -3. We then multiply the first row by -3, and write the result under the second row.

Question 1: why do we multiply the first row by -3? Because that's what we need to cancel the 3 in the second row.

Question 2: why do we write the result under the second row? Because

that's where we're adding the result.

Remember, you must also multiply 1 by -3 and add it to 2. Why? **Equality**: whatever you do to one side of the equation, you **must** do to the other.

[**-3**]

$$\begin{pmatrix} 1 & 2 \\ 3 & 7 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (1) \\ (2) \end{pmatrix}$$

-3 -6 **-3**

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} (1) \\ (-1) \end{pmatrix}$$

Step 2: We can now read off the answers! The two equations are:

$$1x + 2y = 1$$
$$0x + 1y = -1$$

So we learn from the second equation that y = -1. We then substitute that value into the first equation and get 1x + 2(-1) = 1, so x = 3. We can check this by substituting these values for x and y into the original equations:

$$1x + 2y = 1 \rightarrow 1(3) + 2(-1) = 1$$
$$3x + 7y = 2 \rightarrow 3(3) + 7(-1) = 2$$

So yes, these values work.

Let's do a slightly harder example.

Consider the following:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 10 \end{pmatrix}$$

Step 1: we want to get a matrix that has all zeros under the diagonal. So we need to get rid of the 2 in the second row and the 3 in the third row. To get rid of the 2 in the second row, we multiply the first row by **-2** and add the result to the second row; we multiply the first row by **-3** and add the result to the third row. Remember, we write the results of the multiplication under the row we're going to add it to, and remember we **MUST** also do the multiplication on the right hand side. So we must multiply 1 by **-2** and we must multiply 10 by **-3**.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 0 \\ \mathbf{-2} & \mathbf{-4} & \mathbf{-6} \\ 3 & 0 & 1 \\ \mathbf{-3} & \mathbf{-6} & \mathbf{-9} \end{pmatrix} \begin{pmatrix} x \\ y \\ \mathbf{-4} \\ z \\ \mathbf{-6} \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ \\ 10 \\ \end{pmatrix}$$

[-2] [-3]

This gives

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -6 \\ 0 & -6 & -8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$$

We're almost there – we now need to get rid of the -6 in the third row. Then we'll have a matrix with all zeros under the main diagonal, and we'll be able to read off the answers.

Step 2: We need to get rid of the -6 in the third row. There's nothing we can multiply the first row by. Why? If we add copies of the first row to the third, we'll lose the 0 which starts off the third row. What we should do is multiply the second row by something and add it to the third, as this way we won't lose the zero. So, we need to find 'a' such that $(-1)a + (-6) = 0$, hence $a = -6$.

$$
\begin{array}{lll}
 & (1 \;\; 2 \;\; 3)(x) & (2) \\
 & (0 \; -1 \; -6)(y) & = \; (-3) \\
[-6] & (0 \; -6 \; -8)(z) & (\; 4) \\
 & \mathbf{0 \quad 6 \quad 36} & \mathbf{18}
\end{array}
$$

This yields

$$
\begin{array}{lll}
 & (1 \;\; 2 \;\; 3)(x) & (\; 2) \\
 & (0 \; -1 \; -6)(y) & = \; (-3) \\
 & (0 \;\; 0 \; 28)(z) & (22)
\end{array}
$$

Step 3: We can now read off the answers! The three equations are

$$
\begin{aligned}
1x + 2y + 3z &= 2 \\
0x - 1y - 6z &= -3 \\
0x + 0y + 28z &= 22
\end{aligned}
$$

So $z = 22/28 = 11/14$
So $-y - 6(11/14) = -3 \rightarrow y = -24/14$
So $x + 2(-24/14) + 3(11/14) = 2 \rightarrow x = 43/14$

Let's check these numbers in the original equations:

$$
\begin{aligned}
1x + 2y + 3z = 2 &\rightarrow 1(43/14) + 2(-24/14) + 3(11/14) = 2 \\
2x + 3y + 0z = 1 &\rightarrow 2(43/14) + 3(-24/14) + 0(11/14) = 1 \\
3x + 0y + 1z = 10 &\rightarrow 3(43/14) + 0(-24/14) + 1(11/14) = 10
\end{aligned}
$$

So we see we do obtain the correct answer! (If it makes you feel better, I got wrong answers the first two times I did the problem – I did the algebra wrong).

[3] INVERTING MATRICES:

We're now ready to use the method of Gaussian Elimination to invert matrices. Let's review how Gaussian Elimination works. We start off with a matrix A and we do row operations to it. This is equivalent to multiplying A by several matrices E_1, E_2, ..., E_n (say).

For simplicity, let's assume it takes 5 steps to Gaussian Eliminate A to the Identity matrix, so $E_5 E_4 E_3 E_2 E_1 A = I$. Then $E_5 E_4 E_3 E_2 E_1 = A^{-1}$, the inverse matrix to A.

To keep track of these steps, we can just form $E_5 E_4 E_3 E_2 E_1 I$, which by the above is A^{-1}.

An example should illustrate.

Let's try to find the inverse to A =

(1 2)
(3 5)

THE GOAL: We will use Gaussian Elimination to get A to the identity matrix (ones on the main diagonal, zeros elsewhere). We will keep track of the Gaussian Elimination by acting on the Identity matrix.

Step 1: Write the matrix A followed by the identity:

(1 2) (1 0)
(3 5) (0 1)

Step 2: We need to eliminate the 3 in the second row, so we must find 'a' such that 1a + 3 = 0. Hence a = -3. So we multiply the first row of A by -3 and add it to the second row. And remember, by **EQUALITY**, we must do the same to the other side, to the Identity.

(1 2) (1 0)
[-3] (3 5) (0 1)
-3 -6 **-3 0**

(1 2) (1 0)
(0 -1) (-3 1)

Step 3: Now, we want to have all 1s along the main diagonal, so we might as well adjust the second row right now. We have a -1, where we want a 1. So we must multiply the second row by -1. Again, must do this to both sides:

(1 2) (1 0)
(0 -1) (-3 1)
0 1 **3 -1**

Hence we get

(1 2) (1 0)
(0 1) (3 -1)

Step 4: Now we need to get rid of the 2 in the first row, so we multiply the second row by -2 and get:

(1 2) (1 0)
0 -2 **-6 2**
(0 1) (3 -1)

and we get

(1 0) (-5 2)
(0 1) (3 -1)

Note: as a check, you can go thru and see that

(-5 2)
(3 -1)

is the inverse to A.

Let's do one more problem. Let's find the inverse for B =

(9 4)
(7 3)

Step 1: Write the matrix B followed by the Identity:

(9 4) (1 0)
(7 3) (0 1)

Step 2: What should we multiply the first row by to get rid of the 7 in the second row? So, find 'a' such that 9a + 7 = 0, or a = -7/9.

        (9 4)   (1 0)
[-7/9] (7 3)   (0 1)
        -7 -28/9   -7/9 0

And we get

(9 4) (1 0)
(0 -1/9) (-7/9 1)

Step 3: We want to end up with the identity matrix on the left. We have -1/9 in the lower diagonal – we need to multiply the second row by -9 to get 1.

        (9 4)   (1 0)
[-9]   (0 -1/9)   (-7/9 1)
        0   1   7   -9

And we get

(9 4) (1 0)
(0 1) (7 -9)

Step 4: We need to get rid of the 4 in the first row, so we multiply the second row by -4 and add it to the first

$$[-4]\quad \begin{matrix}(9 & 4)\\ 0 & -4\end{matrix} \qquad \begin{matrix}(1 & 0)\\ -28 & 36\end{matrix}$$
$$(0\ 1) \qquad (7\ \ -9)$$

And we get

$$\begin{matrix}(9 & 0)\\ (0 & 1)\end{matrix} \qquad \begin{matrix}(-27 & 36)\\ (\ 7 & -9)\end{matrix}$$

Step 5: We need to have the identity on the left. We have a 9 in the upper left corner, so we must multiply the first row by 1/9.

$$[1/9]\quad \begin{matrix}(9 & 0)\\ 1 & 0\end{matrix} \qquad \begin{matrix}(-27 & 36)\\ -3 & 4\end{matrix}$$
$$(0\ 1) \qquad (\ 7\ \ -9)$$

And we get

$$\begin{matrix}(1 & 0)\\ (0 & 1)\end{matrix} \qquad \begin{matrix}(-3 & 4)\\ (\ 7 & -9)\end{matrix}$$

You can check that this is the inverse of B by doing the multiplication.

NOTES ON LINEAR ALGEBRA

CONTENTS:

[4] MATRIX ADDITION
[5] MATRIX NOTATION
[6] TRANSPOSE
[7] SYMMETRIC MATRICES
[8] BASIC FACTS ABOUT MATRICES

[4] MATRIX ADDITION

Let A and B be two matrices. When can we add them, and what is the answer? We ***define*** matrix addition by adding componentwise. For example:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 7 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 9 \\ 5 & 4 \end{pmatrix}$$

Or

$$\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 0 \end{pmatrix} + \begin{pmatrix} 5 & 7 & 1 \\ 2 & 0 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 9 & 6 \\ 5 & 4 & 8 \end{pmatrix}$$

Of course, we've yet to give any motivation as to *why* one would want to define matrix addition by the above. Remember how we introduced matrices as maps from one space to another. For example, consider the matrix

$$\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 0 \end{pmatrix}$$

It has 2 rows and 3 columns. It acts on vectors with three components, and returns something with 2 components. For example:

$$\begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1*3 + 2*2 + 5*1 \\ 3*3 + 4*2 + 0*1 \end{pmatrix} = \begin{pmatrix} 7 \\ 17 \end{pmatrix}$$

So, if we have two matrices A and B acting on the same vector, we can now see why they should have the same number of rows and columns. They should have the same number of
columns because they both act on the same vector. They should have the same number of ***rows*** because they should each take that vector to the same space.

For example, here's an example of what can go wrong when we try to add two matrices of different sizes.

Consider

$$\begin{matrix}(1\ 3\ 2)\ (3) & & (1*3 + 3*2 + 2*1) & & (11) \\ (2\ 4\ 1)\ (2) & = & (2*3 + 4*2 + 1*1) & = & (15) \\ (4\ 5\ 1)\ (1) & & (4*3 + 5*2 + 1*1) & & (23)\end{matrix}$$

Then

$$\begin{matrix}(1\ 2\ 5)\ (3) & & (1\ 3\ 2)\ (3) & & (\ 7) & & (11) \\ (3\ 4\ 0)\ (2) & + & (2\ 4\ 1)\ (2) & = & (17) & + & (15) \\ \quad (1) & & (4\ 5\ 1)\ (1) & & & & (23)\end{matrix}$$

And we have trouble, as the two vectors are different sizes. One lives in the 2dimensional plane, one lives in 3space. There is no way we can write down one matrix to represent the action of the two matrices.

[5] MATRIX NOTATION

When proving a mathematical theorem, it is **not enough** to check it on a couple of matrices. For example:

CLAIM: For any matrix A, A + A is the zero matrix.

FALSE PROOF:

$$\begin{matrix}(0\ 0) & & (0\ 0) & & (0\ 0) \\ (0\ 0) & + & (0\ 0) & = & (0\ 0)\end{matrix}$$

But ***ANY*** other matrix will not work. If you are trying to ***disprove*** a claim, it ***is*** enough to show that, for a specific example, it fails.

Hence

$$(1\ 2) \qquad (1\ 2) \qquad (2\ 4)$$

$$(3\ 4) + (3\ 4) = (6\ 8)$$

So it is very useful in mathematics to handle a large number of matrices all at once. We don't have the time to check each and every matrix individually, as there are infinitely many matrices!

So, we develop shorthand notation. We represent an arbitrary entry of a matrix A by

$$a_{i,j}$$

The 'i' stands for the i^{th} row, the 'j' stands for the j^{th} column. So, a_{12} means the 1^{st} entry in the 2^{nd} row, a_{22} means the 2^{nd} entry in the 2^{nd} row, and so on.

So, we write an arbitrary 2x2 matrix by

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

We write an aribrary 2x3 matrix by

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

We write an arbitrary 3x3 matrix by

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

And we write an arbitrary mxn music (m rows, n columns) by

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}$$

So, to revisit Matrix addition:

$$\begin{matrix}(a_{11}\ a_{12}\ a_{13}) \\ (a_{21}\ a_{22}\ a_{23})\end{matrix} + \begin{matrix}(b_{11}\ b_{12}\ b_{13}) \\ (b_{21}\ b_{22}\ b_{23})\end{matrix} = \begin{matrix}(a_{11}+b_{11}\ \ a_{12}+b_{12}\ \ a_{13}+b_{13}) \\ (a_{21}+b_{21}\ \ a_{22}+b_{22}\ \ a_{23}+b_{23})\end{matrix}$$

Or, in a specific example:

$$\begin{matrix}(1\ 2\ 3) \\ (4\ 5\ 6)\end{matrix} + \begin{matrix}(1\ 0\ 2) \\ (3\ 1\ 0)\end{matrix} + \begin{matrix}(2\ 2\ 5) \\ (7\ 6\ 6)\end{matrix}$$

[6] TRANSPOSE

We now define the transpose of a matrix. For us, the main use will be in studying symmetric matrices, matrices that are equal to their transpose.

We write A^T for the transpose of the matrix A, and we form A^T as follows: the first row of A becomes the first column of A^T; the second row of A becomes the second column of A^T; the third row of A becomes the third column of A^T; ... ; the last row of A becomes the last row of A^T.

So, if A has 3 rows and 5 columns, then A^T has 3 columns and 5 rows (or as we'd normally write it, 5 rows and 3 columns).

Let's do an example:

$$A = \begin{matrix}(0\ 1\ 1) \\ (1\ 2\ 3)\end{matrix} \quad \text{then } A^T = \begin{matrix}(0\ 1) \\ (1\ 2) \\ (1\ 3)\end{matrix}$$

Or

$$A = \begin{matrix}(1\ 2\ 3\ 4) \\ (0\ 0\ 1\ 2) \\ (5\ 4\ 3\ 2)\end{matrix} \quad \text{then } A^T = \begin{matrix}(1\ 0\ 5) \\ (2\ 0\ 4) \\ (3\ 1\ 3) \\ (4\ 2\ 2)\end{matrix}$$

So, for a 2x3 matrix

$$A = \begin{matrix}(a_{11}\ a_{12}\ a_{13}) \\ (a_{21}\ a_{22}\ a_{23})\end{matrix} \quad \text{then } A^T = \begin{matrix}(a_{11}\ a_{21}) \\ (a_{12}\ a_{22}) \\ (a_{13}\ a_{23})\end{matrix}$$

[7] SYMMETRIC MATRICES

Symmetric matrices are very useful in mathematics, physics, and engineering. First, the definition. We say a matrix A is symmetric if it equals it's tranpose, so $A = A^T$. Later we'll briefly mention why they are useful.

The first thing we note is that for a matrix A to be symmetric A must be a square matrix, namely, A must have the same number of rows and columns. Why? If A has m rows and n columns then A^T has n rows and m columns. Since they're equal, they must have the same number of rows (hence m = n) and the same number of columns (hence n = m). We call matrices with the same number of rows and columns **square** matrices.

For example,

(1 2)
(3 4)

even though the above is a square matrix, is not symmetric, as it's tranpose is

(1 3)
(2 4)

However,

(1 5)
(5 1)

is symmetric, as it does equal its tranpose.

THEOREM: Let A a 2x2 matrix. Then A is Symmetric if it's lower left and upper right entries (a_{21} and a_{12}) are the same.

Proof: We write A as [using a,b,c,d instead of a_{11}, ... as it's easier to view]

(a b)
(c d)

Then A^T is

(a c)
(b d)

And $A = A^T$ means

(a b) (a c)
(c d) = (b d)

Since the two matrices are equal, they must be equal componentwise. So the two upper left entries must be the same. This gives a = a, which imposes no new conditions. Let's look at the other entires. The upper right entires must be the same, which imposes the condition

b = c.

The lower left entries must be the same, which imposes the condition c = b (which we already had), and the two lower right entries must be the same, which imposes d = d.

Hence for a 2x2 matrix A to be symmetric we must have b = c, so the matrix looks like

(a b)
(b c)

What about a 3x3 matrix? Assume a 3x3 matrix A equals its transpose:

(a b c) (a d g)
(d e f) = (b e h)
(g h i) (c f i)

This gives nine conditions:

a = a
b = d
c = g these come from looking at the first row of each side of the above.

d = b (already had)
e = e
f = h these come from looking at the second row of each side

g = c (already had)
h = f (already had)
i = i

Hence the most general 3x3 symmetric matrix looks like

(a b c)
(b e f)
(c f i)

We can, of course, continue to do this for 4x4, 5x5, ..., nxn, ... matrices. The main thing to notice is that symmetric matrices are 'nice' with respect to the main diagonal. (Recall the main diagonal is a_{11}, a_{22}, ..., a_{nn}. We see that for a symmetric matrix, the entry in the i^{th} row and j^{th} column is the same as the entry in the j^{th} row and i^{th} column).

THEOREM: $(A + B)^T = A^T + B^T$ (or, the transpose of a sum is the sum of the transposes).

Proof: Let's do a specific case first.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad B = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix}$$

$$\text{Then } A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \qquad B^T = \begin{pmatrix} 3 & 2 \\ 2 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad A^T + B^T = \begin{pmatrix} 4 & 6 \\ 4 & 6 \\ 4 & 6 \end{pmatrix}$$

And we find that

$$A + B = \begin{pmatrix} 4 & 4 & 4 \\ 6 & 6 & 6 \end{pmatrix} \quad \text{and } (A + B)^T = \begin{pmatrix} 4 & 6 \\ 4 & 6 \\ 4 & 6 \end{pmatrix}$$

Hence we see that $(A + B)^T = A^T + B^T$ ***for these two matrices!!!***

Note that the above is NOT a proof – it is merely a verification in this one particular case. Here's a sketch of the proof.

Consider an arbitrary row, say the 2nd. We want to show that $(A + B)^T = A^T + B^T$ are the same. We'll do this by showing that each column on the left hand side equals the corresponding column on the right hand side.

Let's look at the LHSide first. We add the 2nd row of A to the 2nd row of B, and then this sum becomes the 2nd row of A + B. Taking transposes, this gives the 2nd row of $(A + B)^T$.

Now we examine the RHSide. The 2nd column of A^T is the 2nd row of A; the 2nd column of B^T is the 2nd row of B. So the 2nd column of $A^T + B^T$ is the 2nd row of A plus the 2nd row of B.

So, the 2nd row of $(A + B)^T$ equals the 2nd row of $A^T + B^T$. But there is nothing special about 2 – we could do this equally well for ***any*** column, and we see the two sides are in fact equal.

As promised, a few words about why symmetric matrices are useful. First, they're easier to handle then general matrices, as they only need about half as many entries. Once you specify the entries on the main diagonal and above the diagonal, you know all the entries

(as the entries below the diagonal equal the ones above the diagonal). You've seen in your engineering course one example of where symmetric matrices arise. One common example in mathematical physics is with the matrix of second derivatives. For example, consider the matrix where

$$a_{i,j} = \delta f/\delta x_i \delta x_j.$$

Here f is a function of n variables $(x_1, ..., x_n)$, and $\delta f/\delta x_i \delta x_j$ is the partial derivative of f with respect to x_i and x_j. For "good" functions f we have $\delta f/\delta x_i \delta x_j = \delta f/\delta x_j \delta x_i$ (or, it doesn't matter which order you take the derivatives).

[8] BASIC FACTS ABOUT MATRICES

[1] $A + B = B + A$
[2] $x(A + B) = xA + xB$, where x is any number
[3] $(x+y)A = xA + yB$
[4] AB does not always equal BA
[5] $A(BC) = (AB)C$
[6] A(BC) does not always equal (AC)B (for example, consider A = I)
[7] $AA^{-1} = I$, the Identity matrix
[8] $(A^T)^T = A$
[9] $(A + B)^T = A^T + B^T$
[10] $(xA)^T = x A^T$
[11] $(AB)^{-1} = B^{-1} A^{-1}$
[12] $(AB)^T = B^T A^T$
[13] $(A^{-1})^T = (A^T)^{-1}$

Note: we define, for x a real number and A a matrix, xA to be the matrix whose entries are x times those of A.

Example:

$$\begin{matrix} (1\ 2) \\ 2\,(0\ 1) \\ (3\ 4) \end{matrix} = \begin{matrix} (2\ 4) \\ (0\ 2) \\ (6\ 8) \end{matrix}$$

NOTES ON LINEAR ALGEBRA

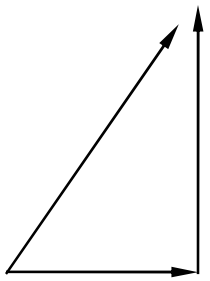
CONTENTS:
[9] BASIS VECTORS

[9] BASIS VECTORS

Consider the vector

$$\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

This means two units in the x direction, five units in the y direction.



Graphically, we see we can write it as a vector in the x direction, and a vector in the y direction. Let

$$Ex = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad Ey = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

be the unit vectors in the x direction and the y direction. We will show that they are a ***basis*** for the plane. What this means is that we can write any vector as some copies of Ex and some copies of Ey.

For example,

$$\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 5 \end{pmatrix} = 2\begin{pmatrix} 0 \\ 1 \end{pmatrix} + 5\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 2\,Ex + 2\,Ey$$

How did we get this? We're trying to write the vector (2,5) as some number of copies of (1,0) and some number of copies of (0,1).

So we're trying to solve

(2) (*) (0)
(5) = (0) + (**)

So, what does * and what does ** equal?

Let's look at the x component, the 'top'. Then 2 = * + 0, so * = 2.
Let's look at the y component, the 'bottom'. Then 5 = 0 + **, so ** = 5.

Let's do another example.

(7) (7) (0) (1) (0)
(3) = (0) + (3) = 7 (0) + 3 (1) = 7 Ex + 3 Ey

Again, let's go through the computation as to how we found it:

(7) (*) (0)
(3) = (0) + (**)

Let's look at the x component, the 'top'. Then 7 = * + 0, so * = 7.
Let's look at the y component, the 'bottom'. Then 3 = 0 + **, so ** = 3.

Now, this leads us to conjecture:

ANY vector in the plane can be written as some number of copies of Ex and some number of copies of Ey.

NOTE: just because we've checked this for several vectors, doesn't mean we've proven the theorem. For example, I might conjecture every 2x2 matrix is symmetric. Why? Well, look at some matrices:

(2 3) (5 0) (5 5) (7 1) (2 0) (2 1) (12 92)
(3 2) (0 5) (5 5) (1 7) (0 5) (1 2) (92 12)

But this is absurd! Consider

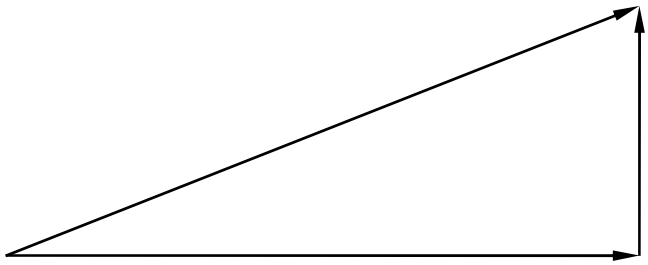
(1 0) (1 2) (5 9)
(2 1) (3 4) (3 3)

So we must be careful not to be misled by checking certain special cases. It is a **very** good idea to test a theorem or conjecture by looking at certain specific cases. This helps lead you to what should be true, but you must prove it in the end.

So, in our case, we must show that, given ***any*** vector (x,y) in the plane, we can find numbers a and b (where a and b will depend on x and y) such that

$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix} = a\,Ex + b\,Ey$$

Now, for the two vectors Ex and Ey, it is easy to find an a and a b. Just take a = x and b = y.



Let's consider a slightly more exotic example.

$$V1 = \begin{pmatrix} 5 \\ 0 \end{pmatrix} \qquad V2 = \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

Are V1 and V2 a basis? Before showing it is, before showing that we can write any vector as copies of V1 plus copies of V2, let's do a specific example first. Consider the vector (1700,-500)

So we want to solve

$$\begin{pmatrix} 1700 \\ -500 \end{pmatrix} = a \begin{pmatrix} 5 \\ 0 \end{pmatrix} + b \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

We are looking for 'a' and 'b'. We have two equations:

$$(Eq1.1) \qquad 1700 = 5a + 12b$$

Unfortunately, this isn't too easy to just look at and see what 'a' and 'b' are. Let's look at the second equation:

$$(Eq1.2) \qquad -500 = 0a + 10b$$

This we ***can*** easily solve. We get 10b = -500, or b = -50. Now that we know b, we can substitute this back into (Eq1.1):

$$1700 = 5a + 12(-50)$$
$$\rightarrow 1700 = 5a - 600$$
$$\rightarrow 2300 = 5a$$
$$\rightarrow a = 460$$

So, we get

$$\begin{pmatrix} 1700 \\ -500 \end{pmatrix} = a \begin{pmatrix} 5 \\ 0 \end{pmatrix} + b \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

or

$$\begin{pmatrix} 1700 \\ -500 \end{pmatrix} = 460 \begin{pmatrix} 5 \\ 0 \end{pmatrix} + -50 \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

So

$$\begin{pmatrix} 1700 \\ -500 \end{pmatrix} = 460\ V1 - 50\ V2$$

Why does this work? Why are V1 and V2 a basis? Notice that while V2 has a piece in the x direction and a piece in the y direction, V1 ***only*** has a piece in the x direction. So if we have a vector (x,y), we must find an a and a b such that (x,y) = aV1 + bV2.

Right now, we don't have to actually ***find*** an a and a b, but just show that we ***could***. We show that we determine b first, and then can find a. Since V2 has a y component, we multiply it by whatever is needed to equal the y component of (x,y). We now have b V2. This has the ***same*** y component as (x,y), but may not have the correct x component.

But this is no problem, as we can still add a number of copies of V1, which is ***only*** in the x direction. So we can add whatever we need to correct the x component.

Now, let's prove that V1 and V2 are a basis. So, given a vector (x,y) we need to find a and b such that (x,y) = a V1 + b V2. Now, a and b will be ***different*** for different values of x and y. Really,

$$a = a(x,y)$$
$$b = b(x,y)$$

So let's find them!

$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} 5 \\ 0 \end{pmatrix} + b \begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

So we must solve

(Eq1.3) $x = 5a + 12b$
(Eq1.4) $y = 0a + 10b$

We can solve (Eq1.4) easily, getting $b = y/10$. Substituting this into (Eq1.3) yields

$$x = 5a + 12b$$
$$x = 5a + 12(y/10)$$
$$x = 5a + 1.2y$$
$$5a = x - 1.2y$$
$$a = (x - 1.2y) / 5$$

So we have succeeded in finding a and b, given x and y!

$$a = (x - 1.2y) / 5$$
$$b = y / 10$$

Note that a and b are different for different values of a and b. For the example we did before, namely (x,y) = (1700,-500) what should a and b be?

Well, the formulas above give

$$a = (1700 - 1.2*(-500)) / 5 = 2300 / 5 = 460$$
$$b = -500 / 5 = -50$$

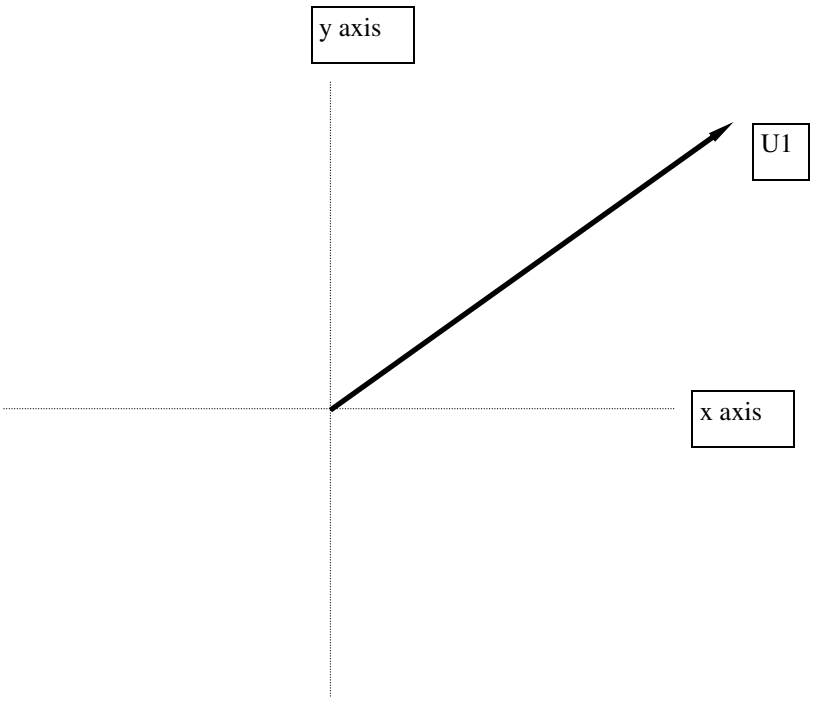
And this agrees with what we calculated before!

From what we've just seen above, we might be led to expect that ***any*** two vectors in the plane are a basis. A quick example shows that there are certain pairs of vectors that cannot be a basis. Consider, for example,

$$U1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad U2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

Now, U1 and U2 are along the same line, as U2 is just twice U1. Any multiple of U1 will be in the same direction as U1; any multiple of U2 will be in the same direction as U2. Hence the sum of a multiple of U1 and a multiple of U2 will ***still*** be in the direction of U1.

Why does this mean U1 and U2 cannot be a basis? Just take ***any*** vector (x,y) that's not in the direction of U1. Then as multiples of U1 and U2 are still in the direction of U1, we cannot get (x,y).



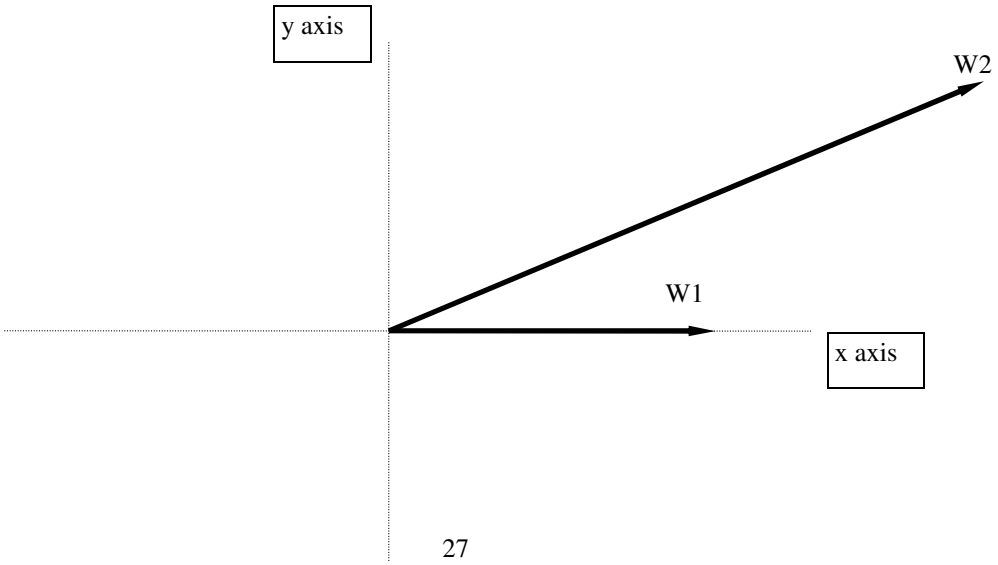
Again, any number of copies of U1 will still be in the direction of U1. If U2 is in the same direction as U1, then copies of U1 plus copies of U2 will ***still*** be in the direction of U1.

So, this quick sketch shows why certain pairs cannot be a basis.

We have the following:

THEOREM: Let W1 and W2 be any two vectors that are not in the same direction (ie, that do not lie on the same line). Then W1 and W2 are a basis.

Let's assume one of the vectors is in the direction of the x axis, and draw a picture.



I don't really want to go into a theoretical, rigorous proof, so I'll just do it in the case when W1 is in the direction of the x axis.

Let's just do a sketch. (Sorry for the pun). The first vector W1 equals, say, $(W1_X, 0)$.

W2 has a non-zero component in the y direction. We're trying to get the vector (x,y).

Let's say W2 = $(W2_X , W2_Y)$. We need to solve

$$(x,y) = a\,W1 + b\,W2$$

$$(x,y) = a\,W1 + b\,(W2_X , W2_Y)$$

If we take $b = y / W2_Y$, (we can divide by $W2_Y$ as it is not zero) then we get

$$(x,y) = a\,W1 + y / W2_Y\,(W2_X , W2_Y)$$

$$(x,y) = a\,W1 + (y\,W2_X / W2_Y, y\,W2_Y / W2_Y)$$

$$(x,y) = a\,(W1_X, 0) + (y\,W2_X / W2_Y, y)$$

$$(x,y) = (a\,W1_X, 0) + (y\,W2_X / W2_Y, y)$$

So the y component on the Left Hand Side equals the y component on the Right Hand Side. We now solve for a:

$$(a\,W1_X, 0) = (x,y) - (y\,W2_X / W2_Y, y)$$

$$(a\,W1_X, 0) = (x - y\,W2_X / W2_Y, 0)$$

Then we can solve for a:

$$a = (x - y\,W2_X / W2_Y) / W1_X.$$

A similar argument would work for ***any*** two vectors W1, W2 that are not in the same direction.

The last thing we're going to do is how to find a and b for W1, W2. I'll give a method (using Gaussian Elimination) that will always work, although I won't prove why.

Let W1 = (A, B) and W2 = (C,D).

We want, given the vector (x,y), to find a and b such that

$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} A \\ B \end{pmatrix} + b \begin{pmatrix} C \\ D \end{pmatrix}$$

So we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} aA \\ aB \end{pmatrix} + \begin{pmatrix} bC \\ bD \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} aA + bC \\ aB + bD \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Aa + Cb \\ Ba + Db \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Now we use Gaussian Elimination to solve! So, we say: what must we multiple the first row (A C) by so that, when we add it to the second row (B D) we get (0 something).

So A*m + B = 0, so we multiply the first row by -B/A. Etc...

NOTES ON LINEAR ALGEBRA

CONTENTS:

[10] BASIS VECTORS - PART II
[11] LINEAR TRANSFORMATIONS

[10] BASIS VECTORS - PART II

We'll now give a procedure to determine when two vectors W1 and W2 are a basis for the plane. Not only will our method say if they're a basis, but it will also tell us how to find a and b.

The Equation we're trying to solve is:

Let

$$\mathbf{W1} = \begin{matrix}\mathbf{(R)}\\ \mathbf{(S)}\end{matrix} \qquad \mathbf{W2} = \begin{matrix}\mathbf{(U)}\\ \mathbf{(V)}\end{matrix}$$

Find a, b so that

$$\begin{matrix}\mathbf{(x)}\\ \mathbf{(y)}\end{matrix} = \mathbf{a}\begin{matrix}\mathbf{(R)}\\ \mathbf{(S)}\end{matrix} + \mathbf{b}\begin{matrix}\mathbf{(U)}\\ \mathbf{(V)}\end{matrix}$$

Then

$$\begin{matrix}\mathbf{(x)}\\ \mathbf{(y)}\end{matrix} = \begin{matrix}\mathbf{(aR)}\\ \mathbf{(aS)}\end{matrix} + \begin{matrix}\mathbf{(bU)}\\ \mathbf{(bV)}\end{matrix}$$

$$\begin{matrix}\mathbf{(x)}\\ \mathbf{(y)}\end{matrix} = \begin{matrix}\mathbf{(Ra)}\\ \mathbf{(Sa)}\end{matrix} + \begin{matrix}\mathbf{(Ub)}\\ \mathbf{(Vb)}\end{matrix}$$

$$\begin{matrix}\mathbf{(x)}\\ \mathbf{(y)}\end{matrix} = \begin{matrix}\mathbf{(Ra + Ub)}\\ \mathbf{(Sa + Vb)}\end{matrix}$$

$$\begin{matrix}\mathbf{(x)}\\ \mathbf{(y)}\end{matrix} = \begin{matrix}\mathbf{(R\ U)\ (a)}\\ \mathbf{(S\ V)\ (b)}\end{matrix}$$

But this equation is just ***begging*** us to use **Gaussian Elimination.** We need to find a number m such that, if we multiply the first row by m and add it to the second, we get the new row will be (0 something).

So $$\mathbf{Rm + S = 0}$$

hence $$\mathbf{m = -S / R}$$

So, we carry out the Gaussian Elimination. It can be shown that if the two vectors (R,S) and (U,V) do not lie on the same line, then Gaussian Elimination will never yield the last row all zero.

Let's do some examples:

FIRST EXAMPLE

$$\mathbf{W1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \qquad W2 = \begin{pmatrix} 3 \\ 7 \end{pmatrix}}$$

Then we must solve

$$\mathbf{\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} 2 \\ 4 \end{pmatrix} + b \begin{pmatrix} 3 \\ 7 \end{pmatrix}}$$

Then

$$\mathbf{\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2a \\ 4a \end{pmatrix} + \begin{pmatrix} 3b \\ 7b \end{pmatrix}}$$

Hence

$$\mathbf{\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 4 & 7 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}}$$

NOW WE DO GAUSSIAN ELIMINATION:

So, what should we multiply the first row by? We need 2m + 4 = 0, so m = -4/2 = -2.

Hence we get

$$\begin{matrix} \mathbf{(x)} & & \mathbf{(2\ 3)} & \mathbf{(a)} \\ \mathbf{(y-2x)} & \mathbf{=} & \mathbf{(0\ 1)} & \mathbf{(b)} \end{matrix}$$

Or $2a + 3b = x$; $0a + 1b = y$

Therefore $b = y$.
$2a + 3b = x \rightarrow a = (x-3b)/2 = (x-3y)/2$

So, given a vector (x,y) we can find a,b such that (x,y) = aW1 + bW2. So W1, W2 is a basis!

SECOND EXAMPLE

$$\mathbf{W1} = \begin{matrix} \mathbf{(2)} \\ \mathbf{(4)} \end{matrix} \qquad \mathbf{W2} = \begin{matrix} \mathbf{(4)} \\ \mathbf{(8)} \end{matrix}$$

Then we must solve

$$\begin{matrix} \mathbf{(x)} & & \mathbf{(2)} & & \mathbf{(4)} \\ \mathbf{(y)} & \mathbf{=} & \mathbf{a\,(4)} & \mathbf{+} & \mathbf{b\,(8)} \end{matrix}$$

Then

$$\begin{matrix} \mathbf{(x)} & & \mathbf{(2a)} & & \mathbf{(4b)} \\ \mathbf{(y)} & \mathbf{=} & \mathbf{(4a)} & \mathbf{+} & \mathbf{(8b)} \end{matrix}$$

Hence

$$\begin{matrix} \mathbf{(x)} & & \mathbf{(2\ 4)} & \mathbf{(a)} \\ \mathbf{(y)} & \mathbf{=} & \mathbf{(4\ 8)} & \mathbf{(b)} \end{matrix}$$

NOW WE DO GAUSSIAN ELIMINATION:

So, what should we multiply the first row by? We need 2m + 4 = 0, so m = -4/2 = -2.

Hence we get

$$\begin{matrix} \mathbf{(x)} & & \mathbf{(2\ 4)\ (a)} \\ \mathbf{(y\text{-}2x)} & \mathbf{=} & \mathbf{(0\ 0)\ (b)} \end{matrix}$$

So, the two equations we must solve are:

$$2a + 4b = x \qquad \text{and} \qquad 0a + 0b = y - 2x$$

Now, regardless of what a and b are, 0a + 0b is ***always*** zero. If y - 2x is not zero, it will be ***impossible*** to solve the second equation. So, can we find x and y such that $y - 2x \neq 0$? Sure. Take x = 0, y non-zero. Or take x nonzero, y = 0. Or take y = 22, x = 12. Almost any choice works.

So we see that W1, W2 are not a basis. We ended up with a row of zeros. Let's look at our two vectors again:

$$\mathbf{W1} = \begin{matrix} \mathbf{(2)} \\ \mathbf{(4)} \end{matrix} \qquad \mathbf{W2} = \begin{matrix} \mathbf{(4)} \\ \mathbf{(8)} \end{matrix}$$

Notice that

$$\mathbf{W2} = \begin{matrix} \mathbf{(2*2)} \\ \mathbf{(2*4)} \end{matrix} = \mathbf{2}\begin{matrix} \mathbf{(2)} \\ \mathbf{(4)} \end{matrix} = \mathbf{2W1}$$

Not only are W1 and W2 not a basis, but they lie on the same line!

THIRD EXAMPLE

$$\mathbf{W1} = \begin{matrix} \mathbf{(1)} \\ \mathbf{(2)} \\ \mathbf{(1)} \end{matrix} \qquad \mathbf{W2} = \begin{matrix} \mathbf{(2)} \\ \mathbf{(2)} \\ \mathbf{(2)} \end{matrix} \qquad \mathbf{W3} = \begin{matrix} \mathbf{(0)} \\ \mathbf{(1)} \\ \mathbf{(1)} \end{matrix}$$

Are W1, W2, W3 a basis?

$$\begin{matrix} \mathbf{(x)} \\ \mathbf{(y)} \\ \mathbf{(z)} \end{matrix} = \mathbf{a}\begin{matrix} \mathbf{(1)} \\ \mathbf{(2)} \\ \mathbf{(1)} \end{matrix} + \mathbf{b}\begin{matrix} \mathbf{(2)} \\ \mathbf{(2)} \\ \mathbf{(2)} \end{matrix} + \mathbf{c}\begin{matrix} \mathbf{(0)} \\ \mathbf{(1)} \\ \mathbf{(1)} \end{matrix}$$

$$\begin{matrix} \mathbf{(x)} \\ \mathbf{(y)} \\ \mathbf{(z)} \end{matrix} = \begin{matrix} \mathbf{(1a)} \\ \mathbf{(2a)} \\ \mathbf{(1a)} \end{matrix} + \begin{matrix} \mathbf{(2b)} \\ \mathbf{(2b)} \\ \mathbf{(2b)} \end{matrix} + \begin{matrix} \mathbf{(0c)} \\ \mathbf{(1c)} \\ \mathbf{(1c)} \end{matrix}$$

(x) **(1a + 2b + 0c)**
(y) = **(2a + 2b + 1c)**
(z) **(1a + 2b + 1c)**

(x) **(1 2 0) (a)**
(y) = **(2 2 1) (b)**
(z) **(1 2 1) (c)**

Now we do Gaussian Elimination! We multiply the first row by -2 and add it to the second row. (1m + 2 = 0, m = -2).

(x) **(1 2 0) (a)**
(y-2x) = **(0 -2 1) (b)**
(z) **(1 2 1) (c)**

Now we multiply the first row by -1 and add it to the third row (1m + 1 = 0, m = -1).

(x) **(1 2 0) (a)**
(y-2x) = **(0 -2 1) (b)**
(z-x) **(0 0 1) (c)**

We don't have to do any more work, as this matrix is ***UPPER TRIANGULAR.*** This means the matrix is all zeros below the main diagonal. We can now solve the three equations, one at a time.

$0a + 0b + 1c = z - x \quad \rightarrow c = z - x$

$0a - 2b + 1c = y - 2x \quad \rightarrow b = (y-2x - z + x) / -2$

$1a + 2b + 0c = x \quad \rightarrow a = y - 2x - z + x$

So these three vectors are a basis.

In general, to determine if something is a basis for 3space:

		(L)			**(P)**			**(S)**
W1	**=**	**(M)**	**W2**	**=**	**(Q)**	**W3**	**=**	**(T)**
		(N)			**(R)**			**(U)**

Are W1, W2, W3 a basis?

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = a\begin{matrix}(L)\\(M)\\(N)\end{matrix} + b\begin{matrix}(P)\\(Q)\\(R)\end{matrix} + c\begin{matrix}(S)\\(T)\\(U)\end{matrix}$$

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = \begin{matrix}(L\,a)\\(M\,a)\\(N\,a)\end{matrix} + \begin{matrix}(P\,b)\\(Q\,b)\\(R\,b)\end{matrix} + \begin{matrix}(S\,c)\\(T\,c)\\(U\,c)\end{matrix}$$

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = \begin{matrix}(L\,a + P\,b + S\,c)\\(M\,a + Q\,b + T\,c)\\(N\,a + R\,b + U\,c)\end{matrix}$$

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = \begin{matrix}(L\ \ P\ \ S)\,(a)\\(M\ Q\ T)\,(b)\\(N\ \ R\ U)\,(c)\end{matrix}$$

The reason for all the colour is to (hopefully) show how things are going. To determine if W1, W2, W3 are a basis, we are led to solving a matrix equation. The first column of our matrix is W1, the second column is W2, the third column is W3. Call this matrix W. We then have

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = a\,W1 + b\,W2 + c\,W3$$

$$\begin{matrix}(x)\\(y)\\(z)\end{matrix} = \begin{matrix}(\qquad\qquad)\,(a)\\(W1\ \ W2\ \ W3)\,(b)\\(\qquad\qquad)\,(c)\end{matrix}$$

[11] LINEAR TRANSFORMATIONS

Linear Transformations are very useful in mathematics. The reason is they allow us to understand functions at complicated values by understanding them at simpler values. First, the definition for functions, then we'll generalize to matrices:

We say a function is a linear function if two conditions hold:

(1) $f(x + y) = f(x) + f(y)$ for all x,y

(2) $f(ax) = af(x)$

Now, it is very unusual for a function to be linear. Take $f(w) = \text{Sin}[w]$.

Then $f(x) = \text{Sin}[ax]$, which usually is not a $\text{Sin}[x]$. For example, if $x = 180$, then a $\text{Sin}[x]$ is always zero. But if $a = ½$, $\text{Sin}[a\,x] = \text{Sin}[90] = 1$.

Let's try $f(w) = w^2$. Does $f(ax) = af(x)$?

Well, $f(ax) = (ax)^2 = a^2 x^2 = a^2 f(x) \neq a f(x)$ unless $a = 1$ or 0.
Also, $f(x+y) = (x+y)^2 = x^2 + 2xy + y^2 = f(x) + 2xy + f(y) \neq f(x) + f(y)$ unless x or y = 0.

How about $f(w) = 3w + 1$?

Well, $f(ax) = 3(ax) + 1$
$= a(3x) + 1$
$= a(3x + 1 - 1) + 1$
$= a(f(x) - 1) + 1$
$= a\,f(x) - a + 1$
$\neq a\,f(x)$ unless $a = 1$

Just in case you're wondering if ***any*** function is linear, here's one that is:

$f(w) = 3w$

Then $f(ax) = 3(ax) = a(3x) = a\,f(x)$

$f(x+y) = 3(x+y) = 3x + 3y = f(x) + f(y)$

[NOTE: one can prove that the only linear functions are $f(x) = cx$, where c is any real or complex number].

We now generalize this to higher dimensions. Why do we care about higher dimensions? Well, matrices act on vectors (you've seen this in your force / stress diagrams) and it turns out that matrices will be linear transformations.

Let **V** and **W** be any two vectors with the same number of components, and let e be a real number. Then any matrix (that is the correct size to act on V and W) is a linear transformation, namely,

(1) $A\,(\mathbf{V} + \mathbf{W}) = A\,\mathbf{V} + A\,\mathbf{W}$
(2) $A(c\,\mathbf{V}) = c\,A\,\mathbf{V}$

I'll sketch the proof for the 2x2 case:

Let **V** = **(v1)** **(v2)** **W** = **(w1)** **(w2)**

Let the matrix **A** = **(a b)** **(c d)**

Then

A (V + W) = **(a b) ((v1) (w1))** **(c d) ((v2) + (w2))**

= **(a b) (v1 + w1)** **(c d) (v2 + w2)**

= **(a(v1 + w1) + b(v2 + w2))** **(c(v1 + w1) + d(v2 + w2))**

= **(av1 + bv2 + aw1 + bw2)** **(cv1 + dv2 + cw1 + dw2)**

= **(av1 + bv2) (aw1 + bw2)** **(cv1 + dv2) + (cw1 + dw2)**

= **(a b) (v1) (a b) (w1)** **(c d) (v2) + (c d) (w2)**

= **A V + A W**

The other condition is even easier to check:

A (eV) = **(a b) ((v1))** **(c d) (** e **(v2))**

= **(a b)** (e **v1)** **(c d)** (e **v2)**

= **(ae v1 + be v2)** **(ce v1 + de v2)**

= e **(a v1 + b v2)** **(c v1 + d v2)**

$$= \quad e \begin{matrix} \mathbf{(a\ b)\ (v1)} \\ \mathbf{(c\ d)\ (v2)} \end{matrix}$$

$$= \quad e\ \mathbf{A\ V}$$

A similar proof works for any size matrix, concluding the proof.

COMING ATTRACTIONS:
WHY DO WE CARE ABOUT BASES? WHY DO WE CARE ABOUT LINEAR TRANSFORMATIONS? WHAT'S THE CONNECTION BETWEEN THE TWO?

Eventually, we'll see that certain matrices have natural 'bases' attached to them. They (and powers of them) may look very ugly as given. But if we changes bases, using something else other than the x-axis and the y-axis, we can often make the matrices look good.

For symmetric matrices, this will be the case. In fact, the **Principle Axis Theorem** says we will be able to find a basis where, if we write our matrix relative to that basis, it will be **diagonal!**

Also, let's say W1 and W2 are a basis. Then we can write any vector V = (x,y) in terms of the two, or

$$V = a\ W1 + b\ W2$$

Then if A is a matrix, we have

$$A\ V = A\ (\ aW1 + bW2\ )$$

$$A\ V = A\ (aW1) + A\ (bW2)$$

$$A\ V = a\ (A\ W1) + b\ (A\ W2)$$

Or, more generally, $A^N\ V = a\ (A^N\ W1) + b\ (A^N\ W2)$

Real Symmetric Matrices have what is called a 'basis of eigenvectors'. That means there are real numbers c_1 and c_2 such that

$$A\ W1 = c_1\ W1 \qquad A\ W2 = c_2\ W2$$

Applying A multiple times yields

$$A^N W1 = c_1{}^N W1 \qquad A^N W2 = c_2{}^N W2$$

Hence

$$A^N V = a (A^N W1) + b (A^N W2)$$

$$= a\, c_1{}^N W1 + b\, c_2{}^N W2 \quad \text{(Eq 11.1)}$$

So here's the advantage: Let's say N is real big, say a billion. If we were to calculate $A^N V$ we would have to multiply A by itself one billion times, and then have that act on V. That's a lot of calculation to do.

But, if our matrix is symmetric, we'll be able to find W1 and W2 (two calculations), c_1 and c_2 (two more calculations) and numbers a and b (two more calculations), and then we just take N = one billion in (Eq 11.1), and we're done!

See how much we saved!

NOTES ON LINEAR ALGEBRA

CONTENTS:
[12] EIGENVALUES / EIGENVECTORS

[12] EIGENVALUES / EIGENVECTORS

In the last section, we examined when two vectors are a basis for the plane. Let's recall what this means. The plane is 2-dimensional. So, we expect that we should be able to specify any vector with two pieces of information, say East component and North component. This corresponds to the standard basis (x-axis, y-axis).

In [11] we saw that, as long as W1 and W2 are not in the same direction, they are a basis for the plane. This means we can write any vector V as V = a W1 + b W2, where a and b are numbers that can be determined, and depend on V.

However, nowhere in [11] did we discuss why we would want to use a basis other than the standard x-axis, y-axis.

The reason is geometry. Often we'll be studying certain matrices that model physical systems. Those physical systems may have certain axes of symmetry, which will often manifest itself in the matrix. And what we will find is that the matrix looks more 'natural', more 'symmetric', if we change basis.

Let A be a matrix acting on a vector v. What can we say about the vector Av?

In general, not much, unless I give you information about A and v. A vector encodes two pieces of information: magnitude, and direction. When we apply a matrix to a vector, we get a new vector. Usually, that vector will have a different magnitude and a different direction.

In terms of computations, this is often unfortunate. We may be interested in some iterative system, where we might have A^{100} v or $A^{6022045}$ v. If Av is in a different direction than v, we have no quick and easy way to calculate A^2 v. Why? We know the magnitude and direction of Av. But we know nothing about A(Av).

If Av is in the same direction as v, however, it's a different ballgame. Let's say Av = 3v. Then we can calculate $A^{\text{any power}}$ v easily.

For example:

$A^2 v = A(Av)$
$= A(3v)$
$= 3(Av)$
$= 3(3v)$
$= 3^2 v$

$A^3 v = A(A^2 v)$
$= A(3^2 v)$ by the previous calculation
$= 3^2(Av)$
$= 3^2(3v)$
$= 3^3 v$

$A^4 v = A(A^3 v)$
$= A(3^3 v)$ by the previous calculation
$= 3^3(Av)$
$= 3^3(3v)$
$= 3^4 v$

$A^5 v = A(A^4 v)$
$= A(3^4 v)$ by the previous calculation
$= 3^4(Av)$
$= 3^4(3v)$
$= 3^5 v$

And a similar argument shows $A^6 = 3^6 v$, ..., $A^n v = 3^n v$.

DEFINITION OF EIGENVALUE / EIGENVECTOR:

We say v is an eigenvector and λ is its eigenvalue if

(1) $A v = \lambda v$

(2) v is not the zero vector

Note that a vector is an eigenvector relative to a given matrix. For example,

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

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

So (1,1) is an eigenvector with eigenvalue 3, while (1,-1) is an eigenvector with eigenvalue 1. But if we consider a different matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

we see (1,1) is not an eigenvector.

Why do we exclude the zero vector? The reason is the zero vector would be an eigenvector for ANY matrix, and ANY number would be its eigenvalue. Let Z be the zero vector. Then

$$A Z = Z = 2 Z = 3 Z = -2342.324 Z$$

and Z would not have a unique eigenvalue. Again, this is a matter of notation / convenience. We will see later its just easier if all eigenvalues are non-zero, for we'll prove certain nice facts about them. For example, a beautiful theorem of Linear Algebra (The Principal Axis Theorem) states that if you have a symmetric real matrix, you can find a basis of mutually perpendicular vectors Vi such that each Vi is an eigenvector of the matrix A! Wow! This means we can compute the action of large powers of symmetric matrices with very little work.

Let's see now how eigenvectors can make life liveable. We'll work with the matrix

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

We saw above that if V = (1,1) and W = (1,-1) then

$$A V = 3 V$$
$$A W = 1 W$$

V and W are not in the same direction, so they're a basis for the plane. So, if you give me any vector (x,y), I can find a and b (depending on x and y) such that

$$(x,y) = a V + b W.$$

Why is this helpful?

Let's build up in stages.

$$A(V + W) = AV + AW$$
$$= 3V + 1W = 3V + W$$

$$A(2V + 11\,W) = A(2V) + A(11\,W)$$
$$= 2\,(A\,V) + 11\,(A\,W)$$
$$= 2\,(3\,V) + 11\,W$$
$$= 6\,V + 11\,W$$

$$A^4\,(2V + 11\,W) = B\,(2\,V + 11\,W), \quad \text{where } B = A^4$$
$$= B\,(2V) + B\,(11\,W)$$
$$= 2\,(B\,V) + 11\,(B\,W)$$
$$= 2\,(A^4\,V) + 11\,(A^4\,W)$$
$$= 2\,(3^4\,V) + 11\,W$$

NOTE: we have the rule

$$A\,(X + Y) = A\,X + AY$$

This is still true for powers of A, as A^4, A^5, etc. are also matrices:

$$A^4\,(X + Y) = A^4\,X + A^4\,Y$$

So, in full generality:

Eq 12.1 $A^n\,(aV + bW)$

$$= A^n\,(aV) + A^n\,(bW)$$
$$= a\,(A^n\,V) + b\,(A^n\,W)$$
$$= a\,(3^n\,V) + b\,W$$

So we see how easy it is to calculate. Let's take n = 2,000,000, and consider A U, where U = (2,3). Now, A U is not in the same direction as U, nor is A^2 U in the same direction as U or A U, etc., so we have 2,000,000 matrix multiplications to do! That's a lot.

The other way is, FIRST, we express U in terms of our nice basis V, W. We do this by Gaussian Elimination:

$$U = a\,V + b\,W$$

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

(2) (1a) (1b)
(3) = (1a) + (-1b)

(2) (1a + 1b)
(3) = (1a + -1b)

(2) (1 1) (a)
(3) = (1 -1) (b)

And then we do Gaussian Elimination. And now we are done! By **Eq 12.1** we are done – all we have to do is put in the values of a and b, and that n = 2,000,000.

Doing the calculation this way is about two steps; the other way is 2,000,000. This is a phenomenal savings. The long way is beyond the strength of the computers -–the numbers are just too large.

Why are we able to have such a savings? This is not a trivial point – it's one of the strengths of Linear Algebra. Linear Algebra is an efficient way of arranging computations. The long way involves lots of hidden cancellation, cancellation that never survives to the end when we group everything.

An example might help. Consider doing the following addition:

```
 1
+5    - 5
      +6    -6
            +7    -7
                  +8    -8
                        +9    -9
```

If we add horizontally, each row is zero but the first. It doesn't matter how many rows I've got, the final answer is just going to be one. There's only one computation to do.

What if we add vertically?

The we get 6 + 1 +1 + 1 + 1 - 9 = 1. We get the same answer, but it's many more steps. The reason is we add 5 then subtract 5. We add 6 then subtract 6. We add 7 then subtract 7. We add 8 then subtract 8. We add 9 then subtract 9. We keep doing things that cancel, but we don't realize it, and hence have to go thru all the steps.

NOTES ON LINEAR ALGEBRA

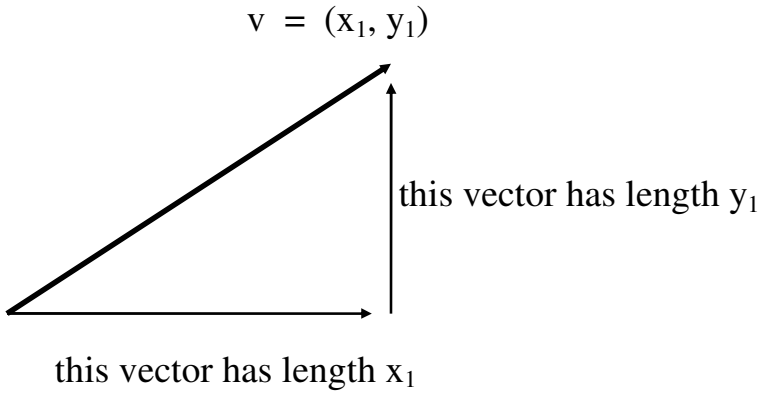
CONTENTS:

[13] DOT PRODUCTS
[14] DETERMINANTS - I

[13] DOT PRODUCTS

The Dot Product is a function from pairs of vectors to numbers. So, the input is two vectors, say $v = (x_1, y_1)$ and $w = (x_2, y_2)$. We use a dot, $\bullet$, to represent the Dot Product.

Let $|v|$ denote the length of the vector v.

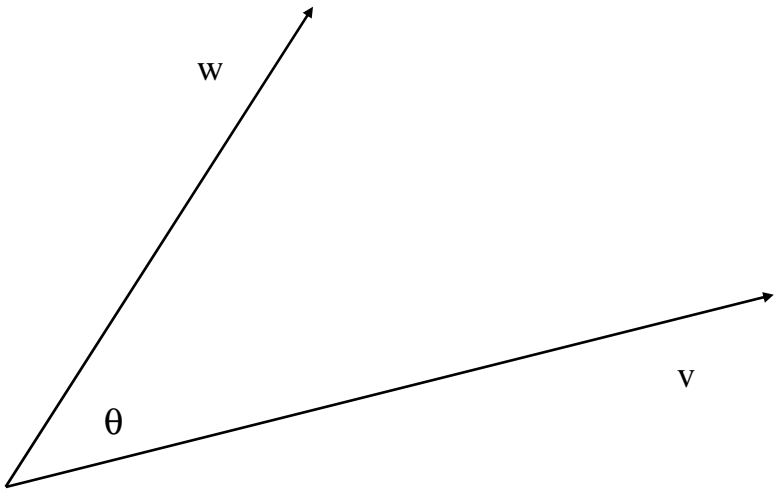


So by the Pythagorean Theorem, the vector v has length $\text{Sqrt}[x_1^2 + y_1^2]$.

Hence $|v| = \sqrt{x_1^2 + y_1^2}$. Similarly $|w| = \sqrt{x_2^2 + y_2^2}$. We now define the Dot Product:

$$v \bullet w = x_1 x_2 + y_1 y_2$$

We will show later that the Dot Product has a very special property, which will explain its usefulness. Namely, if we have



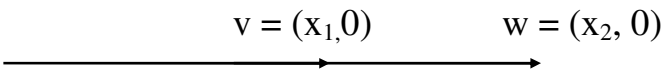
So, if θ is the angle between v and w, then it is a theorem that

$$v \bullet w = |v|\,|w| \cos\theta$$

As always, we will only prove this in two dimensions. Let's look at some special cases. Consider two vectors v and w that are perpendicular, for example:

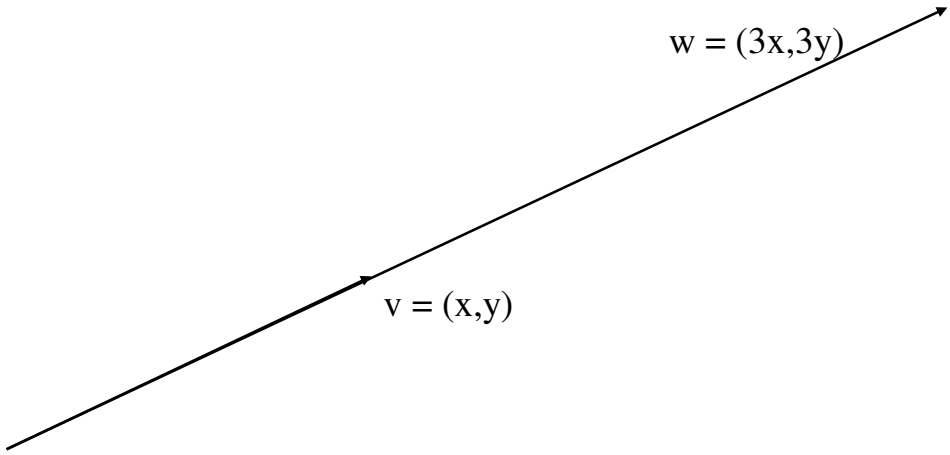
Then $v \bullet w = x_1\,0 + 0\,y_2 = 0$. But as $\theta = 90$, $\cos\theta = 0$, so the formula holds in this case!

Now let's consider v and w in the same direction, say along the x-axis:



Then $v \bullet w = x_1 x_2$. But here $\theta = 0$, so $\cos\theta = 1$, and again the formula works.

Let's do a more exotic example. Let's do v and w in the same direction, but not necessarily along the x-axis.



Now, $|v| = \text{Sqrt}[x^2 + y^2]$, $|w| = \text{Sqrt}[9x^2 + 9y^2]$, $\theta = 0$ so $\cos\theta = 1$.

Then $|v|\,|w| \cos\theta = \text{Sqrt}[x^2 + y^2] * \text{Sqrt}[9x^2 + 9y^2]$
$= \text{Sqrt}[x^2 + y^2] * 3\ \text{Sqrt}[x^2 + y^2]$
$= 3\ (x^2 + y^2)$

And $v \bullet w = x\ 3x + y\ 3y = 3\ (x^2 + y^2)$.

So again, the formula is true.

We now need a linearity property of the Dot Product. Let's say we have three vectors u, v, and w. Then

$$u \bullet (v + w) = u \bullet v + u \bullet w$$

The proof is by straightforward computation. Let's take as our three vectors

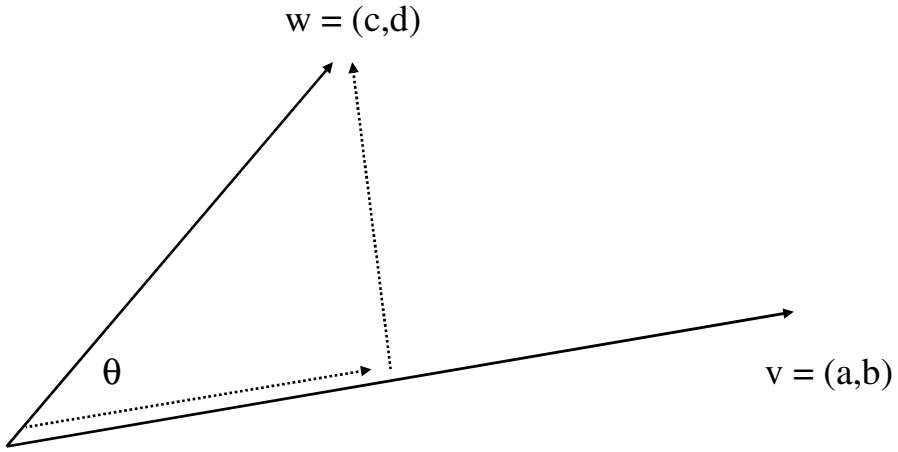
$u = (x_1, y_1)$
$v = (x_2, y_2)$
$w = (x_3, y_3)$

Then $v + w = (x_2 + x_3, y_2 + y_3)$
and $u \bullet (v + w) = x_1 (x_2 + x_3) + y_1 (y_2 + y_3)$
$= x_1 x_2 + x_1 x_3 + y_1 y_2 + y_1 y_3$
$= x_1 x_2 + y_1 y_2 + x_1 x_3 + y_1 y_3$
$= u \bullet v + u \bullet w$

Using all the junk we've just proved, we can now show

$$v \bullet w = |v| |w| \cos\theta$$

Consider two vectors v = (a,b) and w = (c,d):



We break w up into two different vectors:

w_{perp}, which is perpendicular to v, and w_{para}, which is parallel to v.

By the above, we have

$$v \bullet w = v \bullet (w_{perp} + w_{para}) = v \bullet w_{perp} + v \bullet w_{para}$$

But $v \bullet w_{perp} = 0$, and by the special case, $v \bullet w_{para} = |v| |w_{para}| \cos\theta_{v\, w_{para}}$

where $\theta_{v\, w_{para}}$ is the angle between v and w_{para}. But this angle is 0, so we get

$$v \bullet w = v \bullet w_{para} = |v|\,|w_{para}|$$

However, we know what $|w_{para}|$ is – it's just $|w| \cos\theta$. Why? w_{para} is the base of a right triangle with hypotenuse w and angle θ. So substituting above for $|w_{para}|$ yields

$$v \bullet w = |v|\,|w| \cos\theta$$

So we have proved the result in two dimensions. If we were working in 3 space, where we'd have vectors like

$$v = (x_1, y_1, z_1)$$
$$w = (x_2, y_2, z_2)$$

then analogously we define $v \bullet w = x_1 x_2 + y_1 y_2 + z_1 z_2$. Since any two vectors lie in a plane (doesn't matter how many dimensions we are in) we can still talk about the angle between two vectors, and the analogous statement is true.

The three things to take away from Dot Products are:

[1] Two vectors have dot product zero if and only if they are perpendicular

[2] The dot product of two vectors is the product of their lengths if and only if the two vectors are parallel.

[3] The Dot Product measures the angle between two vectors. More precisely, $\cos\theta = v \bullet w / |v|\,|w|$. So, if I know the length of two vectors AND if I know their dot product, I can immediately measure the angle between them!

[14] DETERMINANTS - I

There are several interpretations for Determinant. For now, we will view it as a function whose input is a SQUARE matrix and whose output is a number. We will see in the 2x2 case that this number is the AREA of the parallelogram formed by the rows of A.

If the rows of A are parallel, then this parallelogram will have zero area; if the rows of A aren't parallel, then this parallelogram will have non-zero area. So, the Determinant provides a quick check of whether or not two vectors are in the same direction.

In the plane, this isn't too important; however, in higher dimensions it becomes indispensible. Let's say we are in 3-dimensional space. A is a 3x3 matrix, so its three rows give us three vectors. They form the generalization of a parallelogram, a parallelpiped. (I may have the terminology wrong – it's been a long time since I've used these words!). So instead of talking about area, we should talk about volume. If the three vectors lie in one plane, then this parallelpiped will have zero volume. If the three vectors don't lie in one plane, then the parallelpiped will have non-zero volume. So for 3x3 matrices, the Determinant will measure whether or not the three rows lie in a plane, or if they 'fill' all of three space. Eventually we'll see this is related to questions of when is a matrix invertible.

Now for the definition for 2x2 matrices.

$$\text{Let } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then we denote Determinant(A) several ways:

$$\text{Determinant}(A) = \text{Det}(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Let's see that Det(A) does give the area in certain special cases.

CASE 1:

$$A = \begin{pmatrix} a & b \\ 3a & 3b \end{pmatrix}$$

Then Det(A) = a 3b - b 3a = 0.

Note there's nothing special about 3:

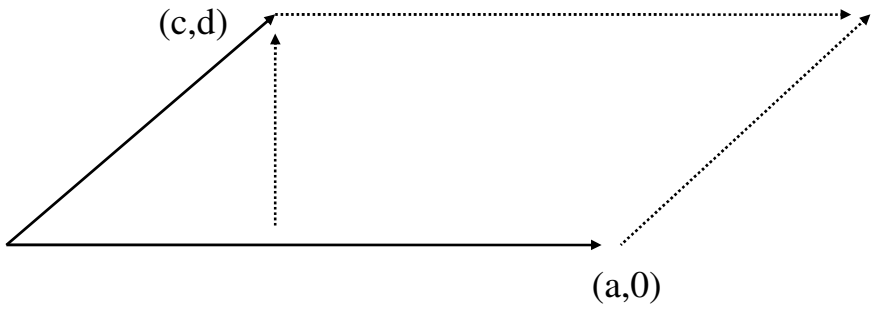
$$A = \begin{pmatrix} a & b \\ ma & mb \end{pmatrix}$$

Then Det(A) = a mb - b ma = 0.

So, when one row is parallel to another, we do get Det(A) = 0!

CASE 2:

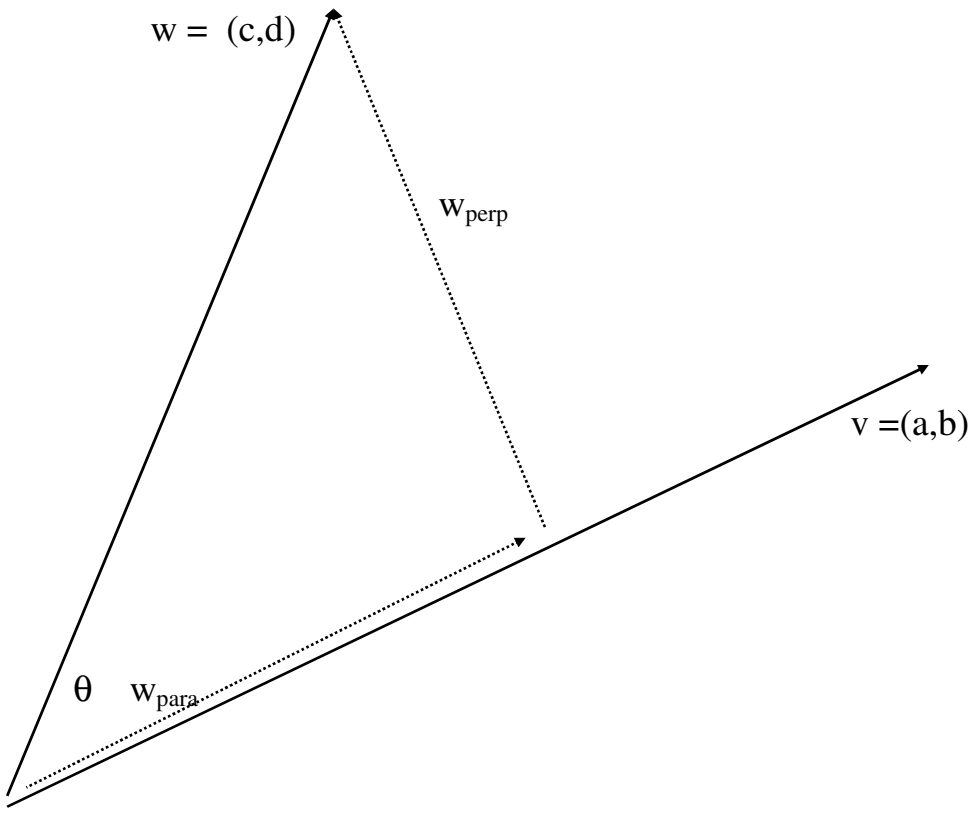
$$A = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$$



The base of the parallelogram is a, the height is d. Hence the area is ad. But Det(A) = ad - 0c = ad. So in this case, it works.

Now we consider the general 2x2 case and, using the Dot Product, we'll prove that Det(A) = area of parallelogram formed by the rows of A.

Again, take A = $\begin{matrix}(a\ b)\\(c\ d)\end{matrix}$



$|v|^2 = a^2 + b^2$ and $|w| = c^2 + d^2$ (by the pythagorean theorem).

$|w_{para}| = |w| \cos\theta$, $|w_{perp}| = |w| \sin\theta$.

So the area of the parallelogram is $|v|\ |w_{perp}|$, or

$$\text{Area} = |v|\ |w_{perp}| = |v|\ |w| \sin\theta$$

But $\cos^2\theta + \sin^2\theta = 1$, so $\sin\theta = \text{Sqrt}[1 - \cos^2\theta]$.

Moreover, $|v|\ |w| \cos\theta = v \bullet w = ac + bd$.

Dividing by $|v|\ |w|$ yields $\cos\theta = (ac+bd) / |v|\ |w|$

Hence Area $= |v|\,|w|\,\mathrm{Sqrt}[1 - \cos^2\theta]$
$= |v|\,|w|\,\mathrm{Sqrt}[1 - (ac+bd)^2 / |v|^2\,|w|^2\,]$
$= \mathrm{Sqrt}[|v|^2\,|w|^2 - (ac+bd)^2]$

Substituting for $|v|^2 = a^2 + b^2$ and $|w| = c^2 + d^2$ yields

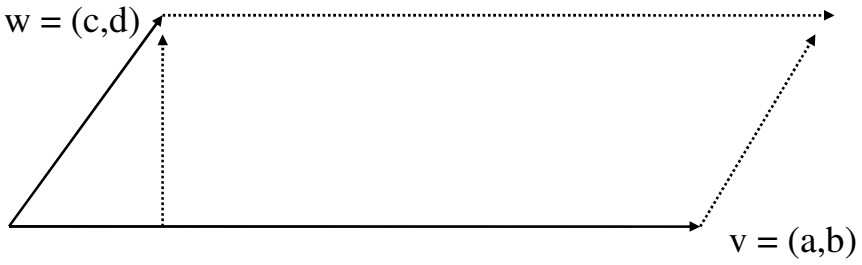
$$\begin{aligned}\text{Area} &= \mathrm{Sqrt}[\,(a^2 + b^2)(\,c^2 + d^2) - (ac+bd)^2\,]\\ &= \mathrm{Sqrt}[\,a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 - a^2c^2 - 2acbd - b^2d^2]\\ &= \mathrm{Sqrt}[a^2d^2 + b^2c^2 \;- 2acbd\,]\\ &= \mathrm{Sqrt}[a^2d^2 - 2adbc + b^2c^2]\\ &= \mathrm{Sqrt}[\,(ad - bc)^2\,]\\ &= ad - bc\end{aligned}$$

So the area of the parallelogram is ad - bc, which is just Det(A)!

One of the reasons Determinant is such a useful function is that say we start with a matrix A, and we do Gaussian Elimination, ending up with a matrix B. Then A and B have the same determinant!

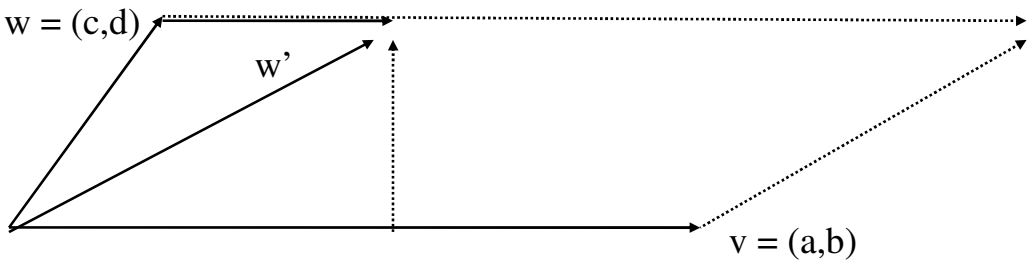
The reason is Gaussian Elimination is just adding multiples of one row to another. So, let's start with the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$



(To simplify things, I'm drawing it as if v is along the x-axis, though the method of proof works in general. This just makes the pictures look nicer).

Now let's say we add on a small multiple of (a,b) to (c,d). So we have a new vector w' = (c+ma, d+mb). Geometrically:



Notice that they have the same base, and the same size height! Hence the two areas are the same.

We can also argue algebraically:

Det(A) = ad - bc.

$$B = \begin{pmatrix} a & b \\ c+ma & d+mb \end{pmatrix}$$

Then Det(B) = a(d+mb) - b(c+ma)
= ad + mab - bc - mab = ad - bc

NOTES ON LINEAR ALGEBRA

CONTENTS:

[15] COMPLEX NUMBERS
[16] FINDING EIGENVALUES

[15] COMPLEX NUMBERS

While we've seen in previous sections how useful eigenvalues and eigenvectors can be, we haven't yet seen how to find them! If it's a very complicated process, then the benefits they provide could be canceled by the work needed to find them. Fortunately, all one needs to do is solve a polynomial and perform Gaussian Elimination.

Somehow, to each square matrix we'll attach a polynomial in one variable, whose degree is the number of columns (or equivalently, the number of rows). So to find the eigenvalues of a 2x2 matrix just requires us to solve a quadratic equation, which is trivial by the quadratic formula.

Unfortunately, a polynomial with real coefficients does not necessarily have real roots. For example, $x^2 + 1 = 0$ has two roots, $x = i$ and $x = -i$, where as always $i = \text{Sqrt}[-1]$.

Reminder: below is a list of types of numbers. Each one is a subset of the next:

[1] Integers: ..., -2, -1, 0, 1, 2, ...
[2] Rationals: p/q, where p, q are integers and $q \neq 0$
[3] Reals: think any terminating or infinite decimal
[4] Complex: of the form a + bi where a and b are real numbers

So, even if we want to study ONLY matrices with real coefficients, we may need to introduce complex numbers to find their eigenvalues. For example, consider the following matrix

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

We'll see later that this has eigenvalues $\pm i$.

However, all is not lost. We have several theorems that will help us in our study:

FUNDAMENTAL THEOREM OF ALGEBRA:
Consider a polynomial of one variable, of degree n. Then there are n roots (not necessarily distinct).

THEOREM OF COMPLEX CONJUGATION:
Let f(x) be a polynomial with real coefficients. Then, if z is a root of f(x) (ie, f(z) = 0), then so is the complex conjugate of z.

EIGENVALUES OF SYMMETRIC MATRICES:
The eigenvalues of symmetric matrices are real if all the entries of the symmetric matrix are real.

Basic properties of complex numbers:

ADDITION:

$$\begin{array}{rr} 2 + 3i & 11 - 7i \\ +\ 4 - 5i & +\ 8 + 8i \\ \hline 6 - 2i & 19 + i \end{array}$$

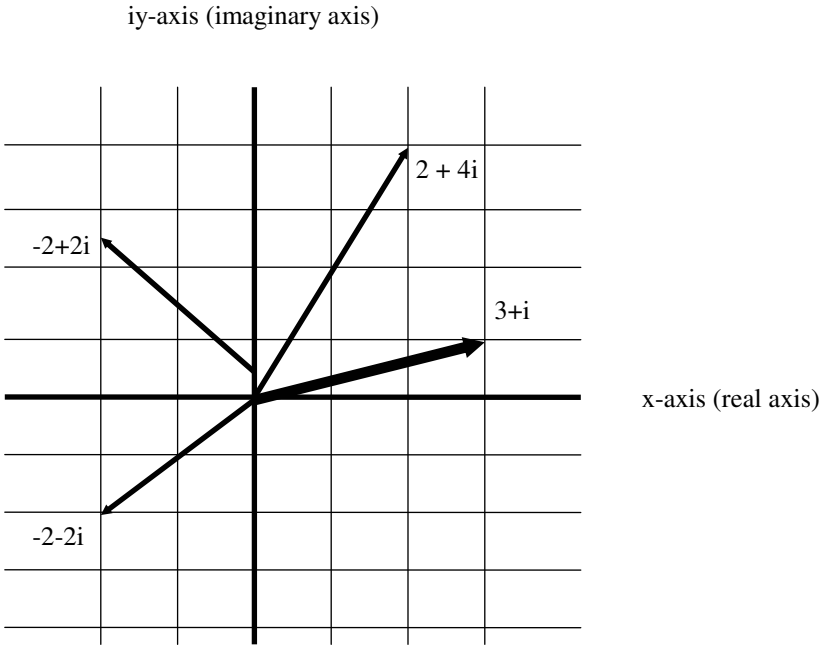
MULTIPLICATION:

Recall (a+b)(c+d) = ac + ad + bc + bd. This is how you multiply complex numbers, but you must remember that $i^2 = -1$.

For example:

$$\begin{aligned}(2 - 3i)(5 + 2i) &= 2*5 + 2*2i - 3i*5 - 3i*2i \\ &= 10 + 4i - 15i - 6i^2 \\ &= 10 - 11i + 6 \\ &= 16 - 11i\end{aligned}$$

GRAPHICAL REPRESENTATION:



COMPLEX CONJUGATION:

If z = x + iy, then $\bar{z}$ = x - iy. We read this the complex conjgate of z.
So 3 - 2i goes to 3 + 2i. -5 - 7i goes to -5 + 7i. -11 goes to -11, 76i goes to -76i. Remember any real number x can be written x + oi. Any number of the form 0 + iy is said to be purely imaginary.

[16a] FINDING EIGENVECTORS (First Version)

A vector has two parts: (1) a direction; (2) a magnitude. Let A be a matrix, and v a vector. Then Av is a new vector. In general, Av and v will be in different directions. However, sometimes one can find a special vector (or vectors) where Av and v have the same direction. In this case we say ***v is an eigenvector of A***. For shorthand, we often drop the 'of A' and say ***v is an eigenvector.***

However, in general v will not equal Av – they may be in the same direction, but they'll differ in magnitude. For example, Av may be twice as long as v, or Av = 2v. Or maybe it's three times, giving Av = 3v. Or maybe it's half as long, and pointing in the opposite direction: Av = -½ v.

In general, we write for v an eigenvector of A:

Av = λ v, where λ is called the eigenvalue.

One caveat: for any matrix A, the zero vector **0** satisfies A **0** = **0.** But it also satisfies A **0** = 2 **0,** A **0** = 3 **0, ...** The zero vector would always be an eigenvector, and any real number would be its eigenvalue. Later you'll see it's useful to have just one eigenvalue for each eigenvector; moreover, you'll also see non-zero eigenvectors encode lots of information about the matrix. Hence we make a definition and require an eigenvector to be non-zero.

The whole point of eigenvectors / eigenvalues is that instead of studying the action of our matrix A on every possible direction, if we can just understand it in a few special ones, we'll understand A completely. Studying the effect of A on the zero vector provides no new information, as EVERY matrix A acting on the zero vector yields the zero vector.

Note: what is an eigenvector for one matrix may not be an eigenvector for another matrix. For example:

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

so here (1,1) is an eigenvector with eigenvalue 3.

(1 1) (1) = (2)
(2 2) (1) (4)

and as (2,4) is not a multiple of (1,1), we see (1,1) is not an eigenvector.

Let's find a method to determine what the eigenvector is, given an eigenvalue.

So, we are given as input a matrix A and an eigenvalue λ, and we are trying to find a non-zero vector v such that $Av = \lambda v$.

Remember, if I is the identity matrix, $Iv = v$ for any vector v. This is the matrix equivalent of multiplying by 1.

$Av = \lambda v$ in algebra, we put all the unknowns on one side. So here we subtract λv from both sides. I'm going to write 0 for the zero vector, but remember, it is not just a number, but a vector where every component is zero.

$Av - \lambda v = 0$

$Av - \lambda Iv = 0$ remember, you're prof is from an Iv-y school: put in the 'I'

$(A - \lambda I) v = 0$

See, λI is a matrix, A is a matrix, so the above is legal. We NEED to put in the Identity matrix. Otherwise, if we went from $Av - \lambda v$ to $(A - \lambda)v$ we'd be in trouble. There, A is a matrix (say 2x2), but λ is a number. And we don't know how to subtract a number from a matrix. We do, however, know how to subtract two matrices. Hence we can calculate what $A - \lambda I$ is, and then do Gaussian Elimination.

Let's do an example:

Say A is

(4 3)
(2 5)

and say $\lambda = 2$. We now try to find the eigenvector v.

$Av = 2v$
$Av - 2v = 0$
$Av - 2Iv = 0$
$(A - 2I)v = 0$

Let's determine the matrix $A - 2I$

I is (1 0) (0 1) so λI is (2 0) (0 2)

Hence

$$A - \lambda I = \begin{pmatrix} 4 & 3 \\ 2 & 5 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix}$$

So we are doing Gaussian elimination on the above matrix. Let's write v = (x,y). Then we must solve:

$$\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So, we multiply the first row by -1 and add it to the second, getting

$$\begin{pmatrix} 2 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The second equation, $0x + 0y = 0$, is satisfied for all x and y. The first equation, $2x + 3y = 0$, says $y = -2/3\ x$. So we see that

$$v = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -2/3\ x \end{pmatrix}$$

Now x is arbitrary, as long as v is not the zero vector. There's many different choices we can make. We can take x = 1 and get the vector (1, -2/3). We can take x = 3, and get the vector v = (3,-2). Notice, however, that the second choice is in the same direction as the first, just a different magnitude.

This reflects the fact that if v is an eigenvector of A, then so is any multiply of v. Moreover, it has the same eigenvalue. Here's the proof:

Say $Av = \lambda v$. Consider the vector $w = mv$, where m is any number.

Then $Aw = A(mv)$
$= m(Av)$
$= m(\lambda v)$
$= \lambda(mv)$
$= \lambda w$

Hence the different choices of x just correspond to taking different lenghts of the eigenvector. Usual choices include x = 1, x = whatever is needed to clear all denominators, and x = whatever is needed to make the vector have length one.

[16b] FINDING EIGENVALUES (Second Version)

We now (finally!) shall see how to find the eigenvalues for a given matrix. Let's look at a **SQUARE** matrix A, and see what numbers can be eigenvalues, and what numbers can't. Let I be the corresponding identity matrix. So, if A is 2x2, I is 2x2; if A is 3x3, I is 3x3, etc.

If λ is an eigenvalue of A, that means there is a non-zero vector v such that

$$A v = \lambda v$$

But Iv = v (as I is the Identity matrix) so

$$A v = \lambda I v$$

$A v - \lambda I v = \mathbf{O}$ where $\mathbf{O}$ is the zero vector.
$(A - \lambda I) v = \mathbf{O}$

Now, $A - \lambda I$ is a new matrix. Let's call it B_λ. Remember how we subtract matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a-e & b-f \\ c-g & h-d \end{pmatrix}$$

So, we are trying to find pairs λ and v (v non-zero) such that

$$B_\lambda v = \mathbf{O}$$

Assume the matrix B_λ is invertible. Then we can multiply both sides by B_λ^{-1} and we get

$$B_\lambda^{-1} B_\lambda v = B_\lambda^{-1} \mathbf{O}$$

But any matrix acting on the zero vector is the zero vector. Hence the Right Hand Side is just $\mathbf{O}$. On the left, $B_\lambda^{-1} B_\lambda = I$, the Identity matrix. So the Left Hand Side is just v.

Hence, if B_λ is invertible, we get $v = \mathbf{O}$. But v must not be the zero vector!

So we have found a necessary condition:

Given a square matrix A, λ is not an eigenvalue of A if $A - \lambda I$ is invertible. Hence the only candidates are those λ such that $A - \lambda I$ is not invertible.

It can actually be shown that this necessary condition is in fact sufficient, namely, if $A - \lambda I$ is not invertible, then λ is an eigenvalue and there is an eigenvector v. Unfortunately, even if the matrix A has all real entries, it's possible that its eigenvector could have complex entries, so we will not give a proof now.

Hence we need an easy way to tell when a matrix is invertible, and when it isn't. It turns out that if A is a square matrix (**remember, only square matrices are invertible**), then A is invertible if and only if Determinant(A) is non-zero. We'll talk more about this later, for now, you may trust Fine Hall.

Given a square matrix A, λ is an eigenvalue of A if and only if Determinant(A - λI) $\neq$ 0.

Let's now do an example. Consider

$$A = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}$$

Now

$$\lambda I = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

and

$$A - \lambda I = \begin{pmatrix} 3-\lambda & 2 \\ 4 & 1-\lambda \end{pmatrix}$$

$$\begin{aligned} \text{Determinant}(A-\lambda I) &= (3-\lambda)(1-\lambda) - (2)(4) \\ &= 3 - \lambda - 3\lambda + \lambda^2 - 8 \\ &= \lambda^2 - 4\lambda - 5 \\ &= (\lambda - 5)(\lambda + 1) \end{aligned}$$

So $\lambda = 5$ or $\lambda = -1$, agreeing with the Homework

Let's do one more example:

$$\begin{pmatrix} 2 & 6 \end{pmatrix}$$

$$A = \begin{pmatrix} 4 & 4 \end{pmatrix}$$

Now

$$\lambda I = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$$

and

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 6 \\ 4 & 4-\lambda \end{pmatrix}$$

$$\begin{aligned} \text{Determinant}(A-\lambda I) &= (2-\lambda)(4-\lambda) - (6)(4) \\ &= 8 - 4\lambda - 2\lambda + \lambda^2 - 24 \\ &= \lambda^2 - 6\lambda - 16 \\ &= (\lambda - 8)(\lambda + 2) \end{aligned}$$

So $\lambda = 8$ or $\lambda = -2$.

NOTES ON LINEAR ALGEBRA

CONTENTS:

[18] VECTORS AND MATRICES
[19] GENERAL REVIEW

[18] VECTORS AND MATRICES

This section will be a general review on the differences between vectors and matrices. Depending on what problem you're studying, there are several different ways of looking at a matrix. For this section, we will look at matrices as maps from one Vector Space to another Vector Space.

We won't go into a technical definition of what a vector space is. Instead, I'll just mention the ones we'll be considering: the set of all vectors with exactly two components; the set of all vectors with exactly three components; the set of all vectors with exactly four components; etc.

Now, a vector has magnitude and direction. Let's take a vector v, and have a matrix A act on it. Now, not every matrix can act on every vector. We have the old row-column rule, which says the number of columns of our Matrix must equal the number of rows of our vector.

Hence

(1 3 5) (4)
(4 6 1) (2)

does not make sense: we get 1*4 + 3*2 + 5*???. However,

(1 3 5) (4)
(4 6 1) (2)
(3 7 9) (3)
(3 4 1)

does make sense.

Let's consider Av, where our matrix A and v are chosen so that this makes sense. For example, we could have

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

and

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Then we find that Av equals

$$\begin{pmatrix} 1x+2y+3z \\ 4x+5y+6z \end{pmatrix}$$

Note that v is in three-space: it has exactly three components. Av, however, is in two-space: it has exactly two components.

Hence we cannot talk about Av + v. It's impossible for v to be an eigenvector for A. Why? Let's say it is an eigenvector, with eigenvalue 2. Then we'd have Av = 2v. The left hand side is a vector with two components. The right hand side is a vector with three components. Trouble!

Think of it as A takes as input vectors with three components, and outputs vectors with just two components. So we cannot talk about Av + v.

This is similar to our troubles with eigenvalue problems. Let's assume now A is a nice 2x2 matrix, say A equals

$$\begin{pmatrix} 5 & 5 \\ 7 & 3 \end{pmatrix}$$

and let's say someone is kind enough to tell us 2 is an eigenvalue, but is unkind enough to ask us to give the corresponding eigenvector. We reason like

$Av = 2v$
$Av - 2v = \mathbf{0}$

But we cannot write (A-2)v. Why? A is a 2x2 matrix, whereas 2 is just a number. Hence A - 2 is not defined. What we can do is remember your professor came from an IVy League school.

For any vector v, Iv = v. So 2Iv = 2v.

IMPORTANT NOTE: we are not saying that 2 = 2I. 2 is a number, 2I is a matrix. What we are saying is that the affect of acting on a vector v with the number 2 is the same as the affect of acting on the vector v by the matrix 2I.

Then we get

$Av - 2Iv = \mathbf{0}$
$(A-2I)\,v = \mathbf{0}$

Let's now quickly review adding vectors. For ease of writing, I'm going to write the vectors horizontally instead of vertically.

So, instead of writing

(1)
(4)

I'll write (1,4).

Let's look at 2(1,4). What does this mean? It means we add two copies of (1,4). The answer is (1,4) + (1,4) = (2,8). One adds vectors by adding them componentwise. So, to add two vectors, they must have the same number of components.

3(1,4) = (1,4) + (1,4) + (1,4) = (1+1+1,4+4+4) = (3,12).

More generally, let r be any number. Then

r(1,4) = (1*r, 4*r).

Fractions get a little tricky, but if you remember the above, it should hopefully lessen the confusion. Let's take, for example,

9/5 (1,4)

Now, if I write 9/5 as 1.8, then it would be

1.8 (1,4) = (1*1.8, 4*1.8) = (1.8, 7.2) = (9/5, 36/5).

When we have a fraction, you just have to remember it's the fraction times the first component is the new first component; the fraction times the second component is the new second component; etc.

So 9/5 (1,4) = (1 * 9/5, 4 * 9/5) = (9/5, 36/5).

NOT 9/5 (1,4) = (1 * 9, 4 * 5). WRONG!

[19] GENERAL REVIEW

This will be a general review on the differences between matrices, vectors, and numbers. Lots of things that you can do with numbers sadly don't hold for matrices. However, some things are the same, so it can get a little confusing. Remember, whenever you write something, you need to have a reason justifying it. Being true for numbers is NOT a valid reason.

Let's look at some examples ***that are true for numbers and matrices:***

1/Addition
3 + 5 = 5 + 3

Or, it doesn't matter what order you add two numbers.

A + B = B + A

For example,

(1 2)	+	(3 4)	=	(4 6)	=	(3 4)	+	(1 2)
(5 6)		(0 1)		(5 7)		(0 1)		(5 6)

So, you can add two matrices in any order.

You can also add two vectors in any order.

2/Multiplying in a Sequence

Recall what 2 * (3 * 4) means. It means FIRST we multiply 3 and 4, THEN we multiply that by 2 on the left. This is the same as (2 * 3) * 4, which means first multiplying 2 by 3, then multiplying that by 4.

For matrices, it's the same. A(BC) = (AB)C. However, please not that we do not have A(BC) = (AC)B. And we also don't have A(BC) = A(CB). We have to keep the matrices in the same order.

Let's look at some things that are different:

3/Getting Zero:

If m and n are two numbers, and mn = 0, then either m = 0, n = 0, or both m and n equal zero. This is not true for multiplying matrices. For example:

Consider the following product:

(0 1) (1 0)
(0 0) (0 0)

How do we find the first column of the product? It's just

(0 1) (1)	=	(0*1 + 1*0)	=	(0)
(0 0) (0)		(0*1 + 0*0)	=	(0)

How do we find the second column? It's just

(0 1) (0) = (0*0 + 1*0) = (0)
(0 0) (0) (0*0 + 0*0) = (0)

Hence

(0 1) (1 0) = (0 0)
(0 0) (0 0) (0 0)

So, even though neither matrix is zero, their product is.

4/Switching Order

For numbers, mn = nm – it doesn't matter which order you multiply them. This, however, is not true for matrices. In general, AB does not equal BA. Let's do a specific example.

(0 1) (0 0)
(0 0) (1 0)

Then the first column is:

(0 1) (0) = (0*1 + 1*1) = (1)
(0 0) (1) (0*0 + 0*1) (0)

And the second column is:

(0 1) (0) = (0*0 + 1*0) = (0)
(0 0) (0) (0*0 + 0*0) (0)

Hence

(0 1) (0 0) = (1 0)
(0 0) (1 0) (0 0)

Let's see what happens if you multiply them the other way:

(0 0) (0 1)
(1 0) (0 0)

Then the first column is:

(0 0) (0) = (0*0 + 0*0) = (0)
(1 0) (0) (1*0 + 0*0) (0)

And the second column is:

(0 0) (1) = (0*1 + 0*0) = (0)
(1 0) (0) (1*1 + 0*0) (1)

So we find that

(0 0) (0 1) = (0 0)
(1 0) (0 0) (1 0)

So the two products are not equal!

5/Now, let's look at what the actions of different objects on other objects give.

For example, let v be a vector, and consider any number, say 2 for definiteness. Then 2v will also be a vector. It will have the same direction as v, but twice the length. Similarly, -3v will point in the opposite direction from v, and be thrice (is that good Queen's English?) the length.

Now let's consider a matrix acting on a vector, say Av. Then this will be a new vector, and except for special v (depending on the matrix A), the direction of Av will not be the same as v. Now, trivially the magnitude of Av will be a multiple of v (think about it – it has to be true!), but as in general their directions are different, it doesn't help us.

Now let's go back to the eigenvalue problem. Let's say we know A, and someone is kind enough to tell you lamda (if a few weeks, you'll know how to find it yourself). Let's say lamda is 5. Then we're trying to solve

$$Av = 5v$$

We remember our algebra, which says we want all the unknowns on one side, so we subtract 5v, and get

$$Av - 5v = 0\text{-vector}$$

Remember: Iv = v. The Identity matrix doesn't change any vector. Hence

If $Iv = v$
then $5Iv = 5v$

So we can substitute for 5v and we go from

$$Av - 5v = 0\text{-vector}$$

to $Av - 5Iv = 0\text{-vector}$

Now, why did we have to introduce the Identity matrix? We would've loved to have been able to go from

$$Av - 5v \quad \text{to} \quad (A-5)v$$

but alas, we cannot. Why? A is a matrix, 5 is a number, and we cannot subtract a number from a matrix.

Note we're never saying 5 = 5I – the left hand side is a number, the right hand side is a vector. What we are saying is 5v = 5Iv.

Now we get

$$(A - 5I)v = 0.$$

And we can solve this by Gaussian Elimination.

6/Mnemonic for multiplying matrices:
Here's a way to remember how to multiply matrices:

Say A is
(1 2)
(2 3)

and B is
(1 0)
(2 1)

And we want to find AB. Well, let's call the first column of B the vector v; let's call the second column of B the vector u. We know how to multiply Av, and we know how to multiply Au. Matrix multiplication is just

AB = A(v u) = (Av Au)

This gives the rule write down the first matrix, then write down the first column of the second matrix. That multiplication gives the first colum of the product matrix AB.

So, in our case:

$$Av = \begin{matrix}(1\ 2)(1) \\ (2\ 3)(2)\end{matrix} = \begin{matrix}(1*2 + 2*2) \\ (2*1 + 3*2)\end{matrix} = \begin{matrix}(6) \\ (8)\end{matrix}$$

Now we do the first matrix times the second column of B to get the second column of the product matrix AB:

$$Av = \begin{matrix}(1\ 2)(0) \\ (2\ 3)(1)\end{matrix} = \begin{matrix}(1*0 + 2*1) \\ (2*0 + 3*1)\end{matrix} = \begin{matrix}(2) \\ (3)\end{matrix}$$

JORDAN CANONICAL FORM

Steven Miller sjmiller@math.ohio-state.edu

I. Introduction:

We've seen that not every matrix is diagonalizable. For example, consider

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Then direct calculation shows that it is not diagonalizable. Why do we care about diagonalizing matrices? The main reason is ease of computation. If we can write $\mathbf{A = S \Lambda S^{-1}}$, then $\mathbf{A^{1000} = S \Lambda^{1000} S^{-1}}$, and the calculations can be performed very quickly. If we had to multiply 1,000 powers of A, this would be very time consuming. Theoretically, we may not need such a time-saving method, but if we're trying to model any physical system or economic model, we're going to want to run calculations on a computer. And if the matrix is decently sized, very quickly these calculations will cause noticeable time-lags.

Jordan Canonical Form is the answer. The Question? What is the 'nicest' form we can get an arbitrary matrix into. We already know that, to every eigenvalue, there is a corresponding eigenvector. If an nxn matrix has n linearly independent eigenvectors, then it is diagonalizable. Hence,

Theorem 1: ***If an nxn matrix A has n distinct eigenvalues, then A is diagonalizable, and for the diagonalizing matrix S we can take the columns to be the n eigenvectors ($S^{-1} A S = \Lambda$).***

In the proof of the above, we see all we needed was n linearly independent vectors. So we obtain

Theorem 2: ***If an nxn matrix A has n linearly independent eigenvectors, then A is diagonalizable, and for the diagonalizing matrix S we can take the columns to be the n eigenvectors ($S^{-1} A S = \Lambda$).***

Now consider the case of an nxn matrix A that does not have n linearly independent eigenvectors. Then we have

Theorem 3: ***If an nxn matrix does not have n linearly independent eigenvectors, then A is not diagonalizable.***

Proof: Assume A is diagonalizable by the matrix S.
Then $S^{-1} A S = \Lambda$, or $A = S \Lambda S^{-1}$.
The standard basis vectors $e_1, ..., e_n$ are eigenvectors of Λ, and as S is invertible, we get $Se_1, ..., Se_n$ are eigenvectors of A, and these n vectors are linearly independent. (Why?) But this contradicts the fact that A does not have n linearly independent eigenvectors. Contradiction, hence A is not diagonalizable.

So, in studying what can be done to an arbitrary nxn matrix, we need only study matrices that do not have n linearly independent eigenvectors.

Jordan Canonical Form Theorem (JCF):
Let A be an nxn matrix. Then there exists an invertible matrix M such that $M^{-1} A M = J$, where J is a block diagonal matrix, and each block is of the form

$$\begin{pmatrix} \lambda & 1 & & & & \\ & \lambda & 1 & & & \\ & & \lambda & 1 & & \\ & & & \ddots & & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix}$$

Note J^{1000} is much easier to computer than A^{1000}. In fact, there is an explicit formula for J^{1000} if you know the eigenvalues and the sizes of each block.

II. Notation:

Recall that λ is an eigenvalue of A if $\text{Det}(A - \lambda I) = 0$, and v is an eigenvector of A with eigenvalue λ if $Av = \lambda v$. We say v is a generalized eigenvector of A with eigenvalue λ if there is some number N such that $(A-\lambda I)^N v = 0$. Note all eigenvectors are generalized eigenvectors.

For notational convenience, we write **gev** for generalized eigenvector, or **λ-gev** for generalized eigenvector corresponding to λ.

We say the **λ-Eigenspace of A** is the subspace spanned by the eigenvectors of A that have eigenvalue λ. Note that this is a subspace, for if v and w are eigenvectors with eigenvalue λ, then av + bw is an eigenvector with eigenvalue λ.

We define the **λ-Generalized Eigenspace of A** to be the subspace of vectors killed by some power of (A-λI). Again, note that this is a subspace.

III. Needed Theorems:

Fundamental Theorem of Algebra: Any polynomial with complex coefficients of degree n has n complex roots (not necessarily distinct).

Cayley-Hamilton Theorem: Let $p(\lambda) = \mathrm{Det}(A-\lambda I)$ be the characteristic polynomial of A. Let $\lambda_1, ..., \lambda_k$ be the distinct roots of this polynomial, with multiplicities $n_1, ..., n_k$ (so $n_1 + ... + n_k = n$). Then we can factor $p(\lambda)$ as

$$p(\lambda) = (\lambda - \lambda_1)^{n1} (\lambda - \lambda_2)^{n2} * ... * (\lambda - \lambda_k)^{nk},$$

and the matrix A satisfies

$$p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk} = 0,$$

Schur's Lemma (Triangularization Lemma): Let A be an nxn matrix. Then there exists a unitary U such that $U^{-1} A U = T$, where T is an upper triangular matrix.

Proof: construct U by fixing one column at a time.

IV. Reduction to Simpler Cases:

In the rest of this handout, we will always assume A has eigenvalues $\lambda_1, ..., \lambda_k$, with multiplicities $n_1, ..., n_k$ (so $n_1 + ... + n_k = n$). We will show that we

can find n_1 λ-gev, n_2 λ-gev, ..., n_k λ-gev, such that these n vectors are linearly independent (LI). These will then form our matrix M.

So, if we can show that the n generalized eigenvectors are linearly independent, and that each one 'block diagonalizes' where it should, it is enough to study each λ separately.

For example, we'll show it's sufficient to consider $\lambda = 0$. Let λ be an eigenvalue of A. Then if v_j is a generalized eigenvector of A with eigenvalue λ, then v_j is a generalized eigenvector with eigenvalue 0 of $B = A - \lambda I$:

$$A\, v_j = \lambda\, v_j + v_{j-1} \quad \rightarrow \quad B\, v_j = 0\, v_j + v_{j-1}.$$

So, if we can find n_j LinIndep gev for B corresponding to 0, we've found n_j LinIindep gev for A corresponding to λ.

The next simplification is that if we can find n_j LinIndep gev for $U^{-1} B U$, then we've found n_j LinIndep gev for B. The proof is a straightforward calculation: let v_1, ..., v_m be the m LinIndp gev for $U^{-1} B U$; then $U^{-1} v_1$, ..., $U^{-1} v_m$ will be m LinIndep gev for B.

Lemma 4: Let $p(\lambda) = (\lambda - \lambda_1)^{n1} (\lambda - \lambda_2)^{n2} * ... * (\lambda - \lambda_k)^{nk}$ be the char poly of A, so $p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk}$. For $1 \leq i \leq k$, consider $(A - \lambda_i I)$. This matrix has exactly n_i LinIndep generalized eigenvectors with eigenvalue 0, hence A has n_i LinIndep generalized eigenvectors with eigenvalue λ_i.

Proof: For notational simplicity, we'll prove this for $\lambda = \lambda_1$, and let's write m for the multiplicity of λ (so $m = n_1$). Further, by the above arguments we see it is sufficient to consider the case $\lambda = 0$. By the Triangularization Lemma, we can put $B = A - \lambda I$ (which has first eigenvalue = 0) into upper triangular form. What we need from the proof is that if we take the first column of U to be v, where v is an eigenvector of B corresponding to eigenvalue 0, then the first column of $T = U_1^{-1} B U_1$ would be $(0,0, ..., 0)^T$.

The lower (n-1)x(n-1) block of T_n, call it C_{n-1}, is upper triangular, hence the eigenvalues of B appear as the entries on the main diagonal. Hence we can

again apply the triangularization argument to C_{n-1}, and get an (n-1)x(n-1) unitary matrix U_{2b}, such that $U_{2b}^{-1} C_{n-1} U_{2b} = T_{n-1}$ has first column (0,0,....0), and the rest is upper triangular. Hence we can form an nxn unitary matrix U_2

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & & & & \\ 0 & & U_{2b} & & \\ \dots & & & & \\ 0 & & & & \end{pmatrix}$$

Then $U_2^{-1} U_1^{-1} B U_1 U_2 =$

$$\begin{pmatrix} 0 & * & * & \dots & * \\ 0 & 0 & * & \dots & * \\ 0 & 0 & * & \dots & * \\ \dots & \dots & \dots & & \dots \\ 0 & 0 & 0 & \dots & * \end{pmatrix}$$

The net result is that we've now rearranged our matrix so that the first two entries on the main diagonal are zero. By 'triangularizing' like this m times, we can continue so that the upper mxm block is upper triangular, with zeros along the main diagonal, and the remaining entries on the main diagonal are non-zero (as we are assuming the multiplicity was m). Call this matrix T_m. Note there is a unitary U such that $T_m = U^{-1}BU$. *Remember, T_m and B are nxn matrices, not mxm matrices.*

Sublemma 1: At most m vectors can be killed by powers of T_m.
Proof: direct calculation: When we multiply powers of T_m, we still have an upper triangular matrix. The entries on the main diagonal are zero for the first m terms, and then non-zero for the remaining terms (because the multiplicity of the eigenvalue $\lambda = 0$ is exactly m). Hence the vectors e_{m+1}, e_{m+2}, ..., e_n are not killed by powers of T_m, and so powers of T_m can have a nullspace of dimension at most m.

We now show that exactly m vectors are killed by T_m. This follows immediately from

Sublemma 2: Let C be an mxm upper triangular matrix with zeros along the main diagonal. Then C^m is the zero matrix.
Proof: straightforward calculation, left to the reader.

Hence the nullspace of $(T_m)^m$ (and all higher powers) is exactly m, which proves there are m generalized eigenvectors of B with eigenvalue $\lambda = 0$. These vectors are LinIndep: As T_m is upper triagonal with zeros on the main diagonal for the first m entries, T_m has m LinIndep gev $e_1, \ldots, e_m$ with eigenvalue 0. As $B = UT_mU^{-1}$, B has m LinIndep gev $Ue_1, \ldots, Ue_m$ with eigenvalue 0 (show that B cannot have any more LinIndep gev with $\lambda = 0$).

Returning to the proof of Lemma 4, we see that there are exactly n_1 vectors killed by $(A-\lambda_1 I)^{n1}$, ..., n_k vectors killed by $(A-\lambda_k I)^{nk}$.

The only reason we go thru this triagonalizing is to conclude that there are exactly n_i vectors killed by $(A-\lambda_i I)^{ni}$. Try to prove this fact directly!

IV. Appendix: Representation of λ-Generalized Eigenvectors.

We know that if λ is an eigenvalue with multiplicity m, there are m generalized eigenvectors, satisfying $(A - \lambda I)^m v = 0$. We describe a very useful way to write these eigenvectors. Let us assume there are t eigenvectors, say $v_1, \ldots, v_t$. We know there are m λ-gev. Note if v is a λ-gev, so is $(A-\lambda I)v$, $(A-\lambda I)^2 v$, …., $(A-\lambda I)^m v$. Of course, some of these may be the zero vector.

We claim that each eigenvector is the termination of some chain of λ-gev. In particular, we have

$$(A-\lambda I)\, v_{1,a} = v_{1,a-1}$$
$$(A-\lambda I)\, v_{1,a-1} = v_{1,a-2}$$
$$\ldots$$
$$(A-\lambda I)\, v_1 = 0 \qquad \text{where } v_1 = v_{1,1}.$$

and

$$(A-\lambda I)\, v_{2,b} = v_{2,b-1}$$
$$(A-\lambda I)\, v_{2,b-1} = v_{2,b-2}$$
$$\vdots$$
$$(A-\lambda I)\, v_2 = 0 \qquad \text{where } v_2 = v_{2,1}.$$

all the way down to

$$(A-\lambda I)\, v_{t,r} = v_{t,r-1}$$
$$(A-\lambda I)\, v_{t,r-1} = v_{t,r-2}$$
$$\vdots$$
$$(A-\lambda I)\, v_t = 0 \qquad \text{where } v_t = v_{t,1},$$

and $a + b + \ldots + r = m$.

We emphasize that we have not shown that such a sequence of λ-gev exists. Later we shall show how to construct these vectors, and then in Lemma 8 we will prove they are Linearly Independent. For now, we assume their existence (and linear independence), and complete the proof of Jordan Canonical Form.

Let us say a λ-gev is a ***pure-gev*** if it is not an eigenvector. Thus, in the above we have t eigenvectors, and m-t pure-generalized eigenvectors. For notational convenience, we often label the λ-generalized eigenvectors by $v_1, \ldots, v_m$. Thus, for a given j, we have $(A-\lambda I)v_j = 0$ if v_j is an eigenvector, and $(A-\lambda I)v_j = v_{j-1}$ if v_j is a pure-gev.

V. Linear Independence of the λ-Generalized Eigenspaces.

Assume the n_1 gev corresponding to λ_1 are linearly independent amongst themselves, and the same for the n_2 gev corresponding to λ_2, ... We now show that the n gev are linearly independent. This fact complete the proof of Jordan Canonical Form (of course, we still must prove the n_i λ_i-gev are linearly independent).

Assume we have some linear combination of the n gev equaling zero. By LC λ_i-gev we mean a linear combination of the n_i λ_i-gev. (This is just to simplify notation).

Then (LC λ_1-gev) + (LC λ_2-gev) + ... + (LC λ_k-gev) = 0.

We'll show first that the coefficients in the first linear combination are all zero. Recall the characteristic polynomial is

$$p(A) = (A - \lambda_1 I)^{n1} (A - \lambda_2 I)^{n2} * ... * (A - \lambda_k I)^{nk}.$$

Define

$$g_1(A) = (A - \lambda_2 I)^{n2} (A - \lambda_3 I)^{n3} * ... * (A - \lambda_k I)^{nk},$$

$g_1(A)$ kills (LC λ_2-gev), $g_1(A)$ kills (LC λ_3-gev),...., $g_1(A)$ kills (LC λ_k-gev). Why? For example, for the λ_2-gev, they are all killed by $(A-\lambda_2 I)^{n2}$, and hence as $g_1(A)$ contains this factor, they are all killed.

What does $g_1(A)$ do to (LC λ_1-gev)? Again, for notational simplicity we'll write m for n_1, and v_1, ..., v_m for the corresponding m λ_1-gev.

We can look at it factor by factor, as all the different terms $(A - \lambda_i I)$ commute.

Lemma 5: For i > 1, let the v_j's be the gev corresponding to λ_1.
If v_j is a pure-gev, then $(A - \lambda_i I)\, v_j = (\lambda_1 - \lambda_i)\, v_j + v_{j-1}$
If v_j is an eigenvector, then $(A - \lambda_i I)\, v_j = (\lambda_1 - \lambda_i)\, v_j$.

Again, the proof is a calculation: if v_j is a pure-gev,

$$\begin{aligned}(A - \lambda_i I)\, v_j &= (A - \lambda_1 I + \lambda_1 I - \lambda_i I)\, v_j \\ &= (A - \lambda_1 I)\, v_j + (\lambda_1 I - \lambda_i I)\, v_j \\ &= v_{j-1} + (\lambda_1 - \lambda_i)\, v_j\end{aligned}$$

The proof when v_j is an eigenvector is similar.

Now we examine $g_1(A)$ ((LC λ_1-gev) + (LC λ_2-gev) + ... + (LC λ_k-gev)) = 0. Clearly $g_1(A)$ kills the last k-1 linear combinations, and we are left with

$$g_1(A)\ (\text{LC } \lambda_1\text{-gev}) = 0$$

Let's say the LC λ_1-gev = $a_1 v_1 + ... + a_m v_m$. We need to show that all the a_j's are zero. (*Remember we are assuming the v_j's are linear independent – we will prove this fact when we construct the v_j's*). Assume $a_m \neq 0$. From our labeling, v_m is either an eigenvector, or a pure-gev that starts a chain leading to an eigenvector: v_m, $(A - \lambda_i I)\, v_m$, $(A - \lambda_i I)2\, v_m$, …. Note no other chain will contain v_m.

We claim that $g_1(A)$ (LC λ_1-gev) will contain a non-zero multiple of v_m. Why? When each factor $(A - \lambda_i I)$ hits a v_j, one gets back $(\lambda_1 - \lambda_i)\, v_j + v_{j-1}$ if v_j is not an eigenvector, and $(\lambda_1 - \lambda_i)\, v_j$ if v_j is an eigenvector. Regardless, we always get back a non-zero multiple of v_j, as $\lambda_1 ? \lambda_i$.

Hence direct calculation shows the coefficient of v_m in $g_1(A)$ (LC λ_1-gev) is

$$a_m (\lambda_1 - \lambda_2)^{n2} (\lambda_1 - \lambda_3)^{n3} * ... * (\lambda_1 - \lambda_k)^{nk}$$

As we are assuming the different λ's are distinct (remember we grouped the eigenvalues together to have multiplicity), this coefficient is non-zero. As we are assuming v_1 thru v_m are linearly independent, the only way the coefficient of v_m can be zero is if $a_m = 0$. Similar reasoning implies $a_{m-1} = 0$, and so on. Hence we have proved:

Theorem 5: Assuming that the n_i generalized eigenvectors associated to the eigenvalue λ_i are linearly independent (for $1 \leq i \leq n$), then the n generalized eigenvectors are linearly independent. Furthermore, there is an invertible M such that $M^{-1} A M = J$.

The only item not immediately clear is what M is. As an exercise, show that one may take M to be the generalized eigenvectors of A. They must be put in a special order. For example, one may group all the λ_1-gev together, the λ_2-gev together, and so on. For each i, order the λ_i-gev as follows: say there are t eigenvectors which give sequences v_1, …, $v_{1,a}$, v_2, …, $v_{2,b}$,…, v_t,…,$v_{t,r}$. Then this ordering works (exercise).

VI. Finding the λ-gev:

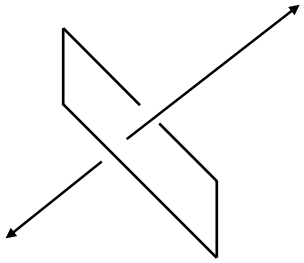
The above arguments show we need only find the n_i generalized eigenvectors corresponding to the eigenvalue λ_i; these will be of the form $(A-\lambda_i I)\, v_j = 0$ or $(A-\lambda_i I)\, v_j = v_{j-1}$. Moreover, we've also seen we may take $\lambda_i = 0$ without loss of generality. For notational convenience, we write λ for λ_i and m for n_i.

So we ***assume the multiplicity of $\lambda = 0$ to be m.*** Hence in the sequel we show how to find m generalized eigenvectors of an mxm matrix whose m^{th} power vanishes. (By the triangularizing we're done, finding m such generalized eigenvectors for this is equivalent to finding m generalized eigenvectors for the original nxn matrix A).

We define the following spaces, where A is our mxm matrix:

1. N(A) = Nullspace(A). The dimension of this is the number of eigenvectors, as we are assuming $\lambda = 0$.
2. $V_1 = W_1 = N(A)$
3. $V_i = N(A^i)$, all vectors killed by A^i. Note that V_m is the entire space.
4. $W_i = \{w \in N(A^i)$ such that $w \perp N(A^{i-1})\}$, for $2 \le i \le m$.

For example, assume we are in R^3, and A^2 is the zero matrix. Let's consider V_2. For definiteness, assume V_1 is 1-dimensional, and V_2 is 3-dimensional. W_1 is just V_1. The problem is, if y_1 and y_2 are two vectors killed by A^2 and not by A, then it is possible that $y_1 - y_2$ (or some linear combination) is killed by A.



In the picture above, the line represents W_1 and the plane represents W_2. Anything in the 3-space above is killed by A^2, and *only* those vectors along the line are killed by just A.. It is possible to take two vectors in R^3 that are linearly independent, neither of which lie on the line, but their difference *does* lie on the line.

Why are we constructing such spaces as W_2? Why isn't V_2 good enough? The reason is we want a very nice basis. The first basis vector will just be a vector in $V_1 = W_1$. For the other two directions, we can take two vectors perpendicular to W_1. (How? This is a 3-dimensional space – simply apply Gram-Schmidt).

The advantage of such a basis is that if z_1 and z_2 are linearly independent vectors in W_2, then the only way a $z_1 + b\ z_2$ can be in W_1 is for $a = b = 0$. Why? W_2 is a subspace, and as z_1 and z_2 are perpendicular to W_1, so is their linear combination. So their linear combination is still in the plane perpendicular to W_1, and as long as a and b are not both zero, it will not be the zero vector in the plane, hence it will be killed by A^2 and not A.

What we are really doing is *Partial Orthogonal Complementation.* Instead of finding the orthogonal complement of V_1 in R^m, we are finding the orthogonal complement in V_2.

Let L be the smallest integer such that A^L is the zero matrix. Then we only need to look up to W_L and V_L. V_L will be an m-dimensional space (as every vector is killed by A^L). We'll have a nice basis for V_L, consisting of the bases of W_1, ..., W_L. The advantage of this decomposition is that the spaces are mutually perpendicular, and if you have a linear combination of vectors in W_j, then the only way it can be in a W_b with $b < j$ is if the combination is the zero vector.

Lemma 6: $\dim(W_{i-1}) \geq \dim(W_i)$, for i = 2, 3, ..., L.
Proof: Assume not: let $N = \dim(W_i)$. So consider a basis of W_i: $z_1, z_2, ..., z_N$, and the vectors $Az_1, ..., Az_N$. Clearly each Az_j is in V_{i-1}. We claim each Az_j must have some component in W_{i-1}. Why? The smallest power that kills each z_j is A^i. If there was no component in W_{i-1}, then A^{i-2} would kill z_j, contradiction.

Let $P = P_{i-1}$ be the projection operator from V_{i-1} to W_{i-1}. Note $P^2 = P$, and by the above arguments each vector $Az_1, ..., Az_N$ has a non-zero component in W_{i-1}. Therefore the N vectors $PAz_1, ..., PAz_N$ are N non-zero vectors in W_{i-1}.

As we are assuming that $\dim(W_{i-1}) < \dim(W_i)$, the N vectors PAz_1 thru PAz_N cannot be linearly independent, for the dimension of a subspace is the

maximal number of linear independent vectors you can have in that space. Hence there exist constants, not all zero, such that

$$a_1 PAz_1 + ... + a_N PAz_N = 0$$

Hence PA $(a_1z_1 + ... + a_Nz_N) = 0$. By the definition of W_i, the linear combination $a_1z_1 + ... + a_Nz_N$ is in W_i. Therefore, the smallest power of A that kills it is I *unless* it is the zero vector. As we are assuming the vectors z_1 through z_N are linearly independent, it is only the zero vector if $a_1 = ... = a_N = 0$.

As $i > 1$, A cannot kill $a_1z_1 + ... + a_Nz_N$ unless this is the zero vector. Could PA kill a non-zero vector? No: by definition, if $a_1z_1 + ... + a_Nz_N$ is not the zero vector, then it is in W_i. Therefore $A(a_1z_1 + ... + a_Nz_N)$ has a non-zero component in W_{i-1} (if not, that contradicts $a_1z_1 + ... + a_Nz_N$ is in W_i). Therefore, $PA(a_1z_1 + ... + a_Nz_N)$ cannot be zero, as $A(a_1z_1 + ... + a_Nz_N)$ has a component in W_{i-1}. Therefore, the only way $PA(a_1z_1 + ... + a_Nz_N)$ can be the zero vector is if $a_1z_1 + ... + a_Nz_N$ is the zero vector, which forces $a_1 = ... = a_N = 0$. Contradiction. QED.

REMARK: by Lemma 6, we see our previous example is impossible. An example that is consistent with Lemma 6 is to consider R^5, let $V_1 = W_1$ be three-dimensional, and W_1 a plane perpendicular to W_1.

Lemma 7: dim(W_i) ≥ 1 for i = 1, 2, ..., L.
Proof: As L is the smallest integer such that A^L is the zero matrix, then there must be a vector killed by A^L but not by A^{L-1}. Whence dim(W_L) is at least 1, so by Lemma 6 we obtain dim(W_i) is at least 1 for i = 1, 2, ..., L.

We now show how to construct the m generalized eigenvectors. We find bases for the spaces W_1, W_2, ..., W_m. We then use A to 'pullback'.

It's easier to explain by an example: assume the dimensions are as follows. Let's take m = 12, and L = 5. For definiteness sake, consider the following:

	$\mathbf{V_1}$	$\mathbf{V_2}$	$\mathbf{V_3}$	$\mathbf{V_4}$	$\mathbf{V_5}$	$\mathbf{V_6}$
	$\mathbf{W_1}$	$\mathbf{W_2}$	$\mathbf{W_3}$	$\mathbf{W_4}$	$\mathbf{W_5}$	$\mathbf{W_6}$
dimW:	4	3	2	2	1	0
basis:	$u_1,...,u_4$	$v_{1,...,}v_3$	$w_{1,}w_2$	x_1,x_2	y	

pullback: $A^4 y \leftarrow A^3 y \leftarrow A^2 y \leftarrow A y \leftarrow y$

Now, W_4 is 2-dimensional.

WE DO NOT KNOW THAT Ay IS IN W_4 !!! IT IS QUITE POSSIBLE THAT Ay IS KILLED BY A^4 AND NO SMALLER POWER OF A WITHOUT BEING IN W_4 !!!

We know y is killed by A^5 and no smaller power of A; hence Ay is killed by A^4 and no smaller power of A. But this does not mean that Ay is in W_4. Fortunately, there is a huge degree of non-uniqueness in the Jordan Canonical Form. We did not need Ay to be in W_4 – all we needed was Ay (and A^2y, A^3y, ...) to be killed by A^4 and nothing lower (A^3 and nothing lower, ...). We'll see below how to handle this.

So for now, all we know is that Ay is in V_4, with a non-zero projection in W_4; that A^2 y is in V_3 with a non-zero projection in W_3, and so on.

W_4 is 2-dimensional. Choose a vector x in W_4 such that x is linearly independent with the projection of Ay onto W_4. Then this x gives us another Jordan Block:

	$\mathbf{V_1}$	$\mathbf{V_2}$	$\mathbf{V_3}$	$\mathbf{V_4}$	$\mathbf{V_5}$	$\mathbf{V_6}$
	$\mathbf{W_1}$	$\mathbf{W_2}$	$\mathbf{W_3}$	$\mathbf{W_4}$	$\mathbf{W_5}$	$\mathbf{W_6}$
dimW:	4	3	2	2	1	0
basis:	$u_1,...,u_4$	$v_{1,...,}v_3$	$w_{1,}w_2$	x_1,x_2	y	

pullback $A^4 y \leftarrow A^3 y \leftarrow A^2 y \leftarrow A y \leftarrow y$
pullback $A^3 x \leftarrow A^2 x \leftarrow A x \leftarrow x$

We continue the game (noting that Ax is in V_3, but not necessarily in W_3). We already have two candidates for directions in V_3, namely A^2y and Ax. We'll show later that though they are not necessarily in W_3, they are killed by A^3 and not A^2, and that their projections onto W_3 are linearly independent.

We need to find 3 directions in W_2. The projections of A^3y and A^2x give us at most two (these two directions could be the same – again, we will show later that this cannot happen). As W_2 is a 3 dimensional space, we can find a vector v in W_3 that is linearly independent with the projections of A^3y and A^2x:

$\mathbf{V_1}$		$\mathbf{V_2}$		$\mathbf{V_3}$		$\mathbf{V_4}$		$\mathbf{V_5}$	$\mathbf{V_6}$
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							

We then have to find four directions in W_1, and have three candidates. We'll see later these three candidates are linearly independent, hence we can find a fourth vector u linearly independent with the rest:

$\mathbf{V_1}$		$\mathbf{V_2}$		$\mathbf{V_3}$		$\mathbf{V_4}$		$\mathbf{V_5}$	$\mathbf{V_6}$
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							
u									

We now have enough (m) candidates. We will show that they are linearly independent.

First we prove that $A^2 y$ and Ax are linearly independent; then we will prove $A^3 y$, $A^2 x$ are linearly independent, and so on. (Proof suggested by L. Fefferman and O. Pascu). Assume $a A^2 y + b Ax = 0$. Then $A(a Ay + b x) = 0$. But Ay has a non-zero projection in W_4, and we've chosen x to be linearly independent in W_4 with the projection of Ay. Therefore the smallest power that can kill this combination is A^4, *unless* it is the zero combination. Hence the only way this can be killed by A is if $a = b = 0$.

Similarly, assume a $A^3 y + b A^2 x = 0$. Then $A^2(a Ay + b x) = 0$, and by the same argument as above, a = b = 0. Note that we did not need Ay to be in W_4, only that it had a non-zero projection there.

By construction, v is linearly independent with $A^3 y$ and $A^2 x$. What about A^4y, A^3x, and Av? Assume $a A^4y + b A^3x + cAv = 0$. Then again we obtain $A(aA^3y + bA^2x + cv) = 0$, and the construction of v forces a = b = c = 0.

Lemma 8: Assume now some linear combination of the m generalized eigenvectors constructed above is zero. Then all the coefficients are zero.

Proof:

V_1		V_2		V_3		V_4		V_5	V_6
$A^4 y$	←	$A^3 y$	←	$A^2 y$	←	$A y$	←	y	
$A^3 x$	←	$A^2 x$	←	$A x$	←	x			
$A v$	←	v							
u									

Assume the coefficient of y, a, is non-zero. Then we have

a y = - (rest of terms), where the rest is killed by A^4, but y isn't.

Hence a must be zero.

Now assume the coefficients of Ay and x are a and b, respectively. Then

a Ay + b x = - (rest of terms), rest killed by A^3.

As x is linearly independent with the projection of Ay onto W_4, a Ay + bx is killed by A^4 and not A^3 *unless* this combination is the zero vectory. As the rest of the terms *are* killed by A^3 , this implies a = b = 0.

Continuing to argue in this way, we obtain all the coefficients are zero, and hence the generalized eigenvectors are linearly independent.

We can now build up our matrix M and J! At last!

For each λ, we have associated generalized eigenvectors. For definiteness sake, let's consider the above case, and I'll leave to you the generalization. We have 4 blocks corresponding to $\lambda = 0$: one block with 5 generalized eigenvectors (starting with the eigenvector A^4y, and ending with y), another block with 4 gev (starting with the eigenvector A^3x, and ending with x), another block of length 2, and one of length 1.

We can order the blocks anyway we want – that will just change the order of the block in J; however, in each block we must write the vectors starting with the eigenvector on the far left, and going to the highest generalized eigenvector on the far right:

$$\begin{pmatrix} A^4y & A^3y & A^2y & Ay & y & A^3x & A^2x & Ax & x & Av & v & u \end{pmatrix}$$

Another possible arrangements would be

$$\begin{pmatrix} A^3x & A^2x & Ax & x & A^4y & A^3y & A^2y & Ay & y & Av & v & u \end{pmatrix}$$

and so on. I'll leave it to you to verify that $M^{-1} A M = J$.

VII. Calculation Shortcut:

When trying to find bases for the spaces W_i, there is a nice shortcut. First, we find a basis for V_i, or start to. How do we find vectors killed by A^i? We just have to find the nullspace of A^i. We do this by Gaussian Elimination, reducing A^i to an upper triangular matrix U. We can assume we've already found bases for W_1 thru W_{i-1}, or equivalently, a basis for V_{i-1}. Let's say the basis for V_{i-1} is $b_1, b_2, ..., b_q$. Then if we add these q rows to U, forming a new matrix U', we observe the following:

(1) If U' v = 0, then v is killed by A^i (from the first m rows of U' are the same as those of U).
(2) If U' v = 0, then v is perpendicular to W_1 thru W_{i-1}: this follows immediately from the fact that we put the basis for W_1 thru W_{i-1} as the last q rows of U', and so this forces v to be perpendicular to these spaces.

Also, if an eigenvalue has multiplicity 3 or less, counting the number of linearly independent eigenvectors gives us the Jordan Form. Why? If there are 3 LI eigenvectors, it's diagonalizable. If there is only 1, it must be a 3x3 block. If there are 2, we must have a 2x2 and a 1x1 block. Note we have no idea what M looks like.

Also note that this argument fails for multiplicity 4 and greater. If we have multiplicity 4 and 2 eigenvectors, it could be 2x2, 2x2, or it could be 3x3, 1x1.

Note the difference between theory and practice: theoretically, we know that bases for the different W_i exist, so with a wave of the hand we have them to work with. But if we were actually going to Jordanize large matrices, finding bases for all these spaces takes time, and we don't always need all those basis elements. Often it's enough to just find vectors in V_i; for example, show that if instead of taking y in W_L we took y in V_L the pullback process would work. Then there could be many i where we've pulled-back all the vectors we need, and hence there would be no need there to find a basis. If this doesn't make too much sense, don't worry: it's late at night here for me, and at this stage in your life, you won't be dealing with terrible Jordanizations where this would really make a difference. I just want to emphasize that often you can come up with a theoretical line of argument that, in practice, will yield the correct answer, but be so computationally inefficient that a better way is greatly desired.

SUMMARY: HOW TO JORDANIZE:

STEP 1: Find the eigenvalues, their multiplicities, and all the eigenvectors.
STEP 2: For each eigenvalue λ and it's multiplicity m, calculate $(A-\lambda I)$, $(A-\lambda I)^2$, ..., $(A-\lambda I)^m$.
STEP 3: Find bases for the spaces W_i described above. This will yield bases for V_i. Use the calculation shortcut to find the bases.
STEP 4: 'Pullback' vectors as described, add in vectors linearly independent with projections as needed.