This file contains the notes I wrote to myself in preparation for the lectures. I make no claims for their completeness or accuracy, but I'm making them available in case they can be of use to anyone.

by Martin Bays mbays@sdf.org

Lecture 1: Linear equations

2x+5y=1

3x-y+\sqrt{2}z=37

ax+by+cz=d

Definition: **A _linear equation_** is an equation of the form
        a_1x_1+a_2x_2+...+a_nx_n=b
    where **a_i** and **b** are constants, and **x_i** are variables.

    // "Sum of constant multiples of variables equals constant".

Geometrically:
    the solutions **(x,y)** to a linear equation **ax+by=c** in two variables form a line in the plane.

    three variables, **ax+by+cz=d**: plane in **R^3**.

    "and so on".

Zero curvature. // Constant derivatives.

Interplay between algebra and geometry;
    geometric intuition, algebraic techniques.

Systems:
    Given a _system_ of linear equations, e.g.
        ax+by+cz=d
        ex+fy+gz=h
    want to "solve the system"
    i.e. describe the solutions
    i.e. determine which **(x,y,z)** satisfy all of the equations

    Geometrically: want to find the intersection of the corresponding planes.

    In this case, we have three possibilities:
        (**i**) intersection is a plane
                e.g. **x+3y+2z=3**
                    2x+6y+4z=6
                equations define same plane

        (ii) intersection is empty
                e.g. **x+3y+2z=3**
                    2x+6y+4z=5
                equations define distinct parallel planes

        (iii) intersection is a line (usual case)
                e.g.
                    x=1
                    y=2
                solutions are **(1,2,t)** where **t** is any number

                or
                    x+y=1
                    x-z=3
                can rewrite as **y=1**-x, **z=x-3**
                solutions are **(t,1-t,t-3)** where **t** is any number

    // With three equations in three variables (intersecting three planes),
    // can have no solutions, a single solution, a line of solutions or a
    // plane of solutions.

Solving systems algebraically:

// We'll give a complete algorithm ("Gaussian elimination") on Wednesday.
// Let's solve some by ad hoc methods now, which will turn out to be those
// involved in Gaussian elimination.

Basic idea: transform the system into a "simpler" system with the same solutions for which it's obvious what the solutions are.

Example 1:

$$2x + 3y = 4$$
$$4x - y = 15$$

For any **c**, if we replace the second equation with the second equation plus **c** times the first equation, forming the system

$$2x + 3y = 4$$
$$(4x - y) + c(2x+3y) = 15 + c4$$

then this new system has the same solutions as the old system:

Suppose **(x,y)** satisfies the original system, then obviously it satisfies the first equation of the new system, and it's easy to check that it also satisfies the second equation.

What's more, we get no new solutions: if **(x,y)** satisfies the new system, then

$$4x - y = (4x - y) + c(2x+3y) - c(2x+3y)$$
$$= 15 + c4 - c4$$
$$= 15$$

so it satisfies the old system too.

In particular, we can take **c** to be −2, and then the new system is

$$2x + 3y = 4$$
$$(4x - y) + (-2)(2x+3y) = 15 + (-2)4$$

collecting terms:

$$2x + 3y = 4$$
$$(4 + (-2)2)x + (-1 + (-2)3)y = 15 + (-2)4$$

adding:

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

multiplying through second equation by **−1/7**:

$$2x + 3y = 4$$
$$y = -1$$

again, this doesn't change the solutions: **−7y=7** if and only if **y=−1.**

now if we add −3 times the second equation to the first, we get

$$2x = 4 + 3$$
$$y = -1$$

so

$$x = 7/2$$
$$y = -1$$

At each stage, we've made sure that the new system has the same solutions as the old one. The last system clearly has just one solution **(7/2,−1)**, so this is also the only solution of the original system.

Example 2:

$$x + y + z = 0$$
$$x - z = 1$$
$$y + 2z = 5$$

Subtract second equation from first (i.e. add −1 times it):

$$y + 2z = -1$$
$$x - z = 1$$
$$y + 2z = 5$$

As in Example 1, this system has the same solutions as the original system. But clearly the new system has ***no*** solutions, since **y+2z** can't be both −1 and 5. So the original system had no solutions.

Example 3:

$$x + y + z = 0$$
$$x - z = 1$$
$$2y + 4z = -2$$

Again, subtracting second from first yields:

```
          y + 2z = -1
     x      -  z = 1
         2y + 4z = -2
```

Now subtracting twice the first from the third yields:

```
          y + 2z = -1
     x      -  z = 1
               0 = 0
```

so now we just have two equations, which we can rewrite as:

```
     x = 1+z
     y = -1-2z
```

so the solutions form the line (1+t, -1-2t, t)

Lecture 2: Gaussian elimination

We present a systematic method for solving systems of linear equations, using the ideas introduced in lecture 1.

Represent a system

```
    ax+by+cz=d
    ex+fy+gz=h
```

by the "augmented matrix"

```
    ( a b c|d )
    ( e f g|h )
```

(the vertical line is optional)

// We then transform this matrix into one of a simple form - "reduced row echelon
// form", in such a way that the corresponding system will have the same
// solutions as the original system.

// Once in the simple form, we'll be able to "read off" a description of the
// solutions.

An augmented matrix is in _reduced row echelon form_ (rref) if it looks like something like this:

```
    ( 0 1 * * 0 * 0|* )
    ( 0 0 0 0 1 * 0|* )
    ( 0 0 0 0 0 0 1|* )
    ( 0 0 0 0 0 0 0|0 )
```

where the stars can be any numbers (including 0 and 1).

Precise definition of rref:
- (i) Each row either consists of zeros or has first non-zero entry 1 - this is the "leading 1" of the row.
- (ii) Each row after the first starts with at least as many zeros as the previous row does.
- (iii) A column containing a leading 1 has zeros everywhere else.

// note the partitioning line plays no rôle here

/*
An augmented matrix is in _row echelon form_ (ref) if it satisfies (i) and (ii) and (iii'):
    (iii') A column containing a leading 1 has zeros below the 1.
*/

From Webster's Revised Unabridged Dictionary (1913) [web1913]:

  Echelon \Ech"e*lon\ ([e^]sh"e*l[o^]n), n. [F., fr. ['e]chelle
     ladder, fr. L. scala.]

1. (Mil.) An arrangement of a body of troops when its divisions are drawn up in parallel lines each to the right or the left of the one in advance of it, like the steps of a ladder in position for climbing. Also used adjectively; as, echelon distance. --Upton (Tactics).

We use _elementary row operations_ to transform an augmented matrix into rref:

(I) Multiply a row by a non-zero constant
(II) Swap two rows
(III) Add a constant multiple of a row to a different row

Example:

```
    2x + 3y = 4
    4x -  y = 15
```

Corresponding augmented matrix:

```
    ( 2  3|4 )
    ( 4 -1|15 )
```

Multiply first row by 1/2:

```
    ( 1  3/2|2 )
    ( 4  -1 |15 )
```

Add -4 times first row to second row:

```
    ( 1  3/2|2 )
    ( 0  -7 |7 )
```

Multiply second row by -1/7:

```
    ( 1  3/2|2   )
    ( 0   1 |-1 )
```

Add -3/2 times second row to first row:

```
    ( 1   0 |7/2)
    ( 0   1 |-1 )
```

This is in rref. The corresponding system of linear equations is:

x = 7/2
y = -1

The point:

If we apply an elementary row operation to the augmented matrix of a system of linear equations, then the solutions to the system of equations corresponding to the resulting augmented matrix will be the same as those of the original system.

This is obvious for (I) and (II), and almost obvious for (III) (and we discussed it in an example last lecture).

Describing the solutions to a system of equations corresponding to a rref augmented matrix:

If we have a line of the form

```
    ( 0 0 0 0 | 1 )
```

corresponding to an equation

0=1

then there are no solutions.

Otherwise, if we pick arbitrary values for the variables corresponding to columns which do not contain a leading 1, then the values of the other variables are uniquely determined.

e.g.

```
( 0 1 5 0 |27 )
( 0 0 0 1 |1  )
( 0 0 0 0 |0  )
```

corresponds to system

y + 5z = 27
w = 1
0=0

If we set **x** and **z** to arbitrary values, **x=t** and **z=s**, then we read off
y = 27 - 5s
w = 1
so we get a solution **(t,27-5s,s,1)**, and every solution looks like that.

We call this a _parametric_ description of the solutions to the system; here the parameters are **s** and **t**.

Gaussian elimination:
Any augmented matrix can be transformed by a succession of elementary row operations to an rref augmented matrix:

[Do example simultaneously; say

```
( 0 0 1 | 1 )
( 2 0 8 | 4 )
( 1 0 1 | -1 )

( 1 0 4 | 2 )
( 0 0 1 | 1 )
( 0 0 -3 | -3 )

( 1 0 4 | 2 )
( 0 0 1 | 1 )
( 0 0 0 | 0 )
```

Stage I:
Step 1. Swapping rows as necessary, make sure the leftmost non-zero column doesn't start with 0. If there are no non-zero columns, move on to stage II.
Step 2. Divide the first row through so it has a leading 1
Step 3. Add multiples of the first row to the lower rows, such that the leading 1 of the first row has zeros below.
Step 4. Cover up the first row, and repeat from 1 with the resulting shorter matrix.
Stage II:
After stage I, the matrix satisfies (**i**) and (ii) of the definition of rref. The entries _below_ any leading 1 are 0, but those above may be non-zero.

(Such an augmented matrix is said to be in _row echelon **form_**.)

Add multiples of the last non-zero row to previous rows to ensure there are 0s above the leading 1 of the row. Repeat with the second-to-last row, and so on moving up.

We end up with an rref augmented matrix.

Remark:
The textbook refers to this as "Gauss-Jordan" elimination.

Example: solve some systems

```
(1 1  1 |  0)
(1 0 -1 |  1)
(0 1  2 | -2)

(1 2 | 3)
(2 4 | 7)
```

Lecture 3: Matrix algebra

// which in reality I'm doing in lecture 4... and should preface with a quick
// example of Stage 2 of Gaussian elimination, with nice numbers, e.g.
// **(1 2 3 4 | 5)**
// **(0 1 2 3 | 4)**
// **(0 0 0 1 | 2)**
// and make remark about "Gauss-Jordan":
// Remark: The procedure we have described is technically known as
// "Gauss-Jordan elimination"; Stage I on its own (getting a ref matrix) is

// known as "Gaussian elimination".

A _matrix_ is a rectangular array of numbers.
If it has **m** rows and **n** columns, we call it an "**m x n** matrix".
// e.g. blah

An augmented matrix is a matrix - ignore the vertical line.

n x n matrices are _square_ matrices.

Matrices are commonly denoted by capital letters - e.g. **A**, **B**, **C**, **M**...

If **A** is a matrix, we write (**A)_ij** or just **A_ij** for the entry in the ith row and jth column.

Given numbers **a_11,a_12,...,a_1n,a_21,a_22,...,a_2n,...,a_m1,a_m2,...,a_mn**
, we write **[a_ij]** for the matrix **A** such that (**A)_ij = a_ij.**

Two matrices **A** and **B** are equal iff they are of the same size and all their entries are equal: **A_ij = B_ij** for all **i** and **j.**

A _column vector_ is an **m x 1** matrix. **A _row vector_** is a 1 **x n** matrix.
Row and column vectors are commonly denoted by lower case bold or underlined letters. If **b** is a row vector, **b_i** is the entry in the ith column. Similarly for column vectors.

Matrix addition:
    If **A** and **B** are **m x n** matrices, then (**A+B**) is the **m x n** matrix with entries:
        (**A+B)_ij = A_ij + B_ij**

Scalar multiplication of matrices:
    if **c** is a number and **A** is an **m x n** matrix, then cA is the **m x n** matrix with entries:
        (**cA)_ij = cA_ij**

Matrix multiplication:
    If **A** is an **m x n** matrix and **b** is an **n x 1** column vector,
    then we define Ab to be the **m x 1** column vector with entries
        (**Ab)_i = A_i1b_1 + ... + A_inb_n .**

    Why do we define it this way?

    If we have a system of linear equations
        ax **+** by **+** cz **= d**
        ex **+** fy **+** gz **= h**

    we could just as well write it
        a_11 x_1 + a_12 x_2 + a_13 x_3 = c_1
        a_21 x_1 + a_22 x_2 + a_23 x_3 = c_2

    And then if we let **A = [a_ij]**, **x = (x_1,x_2,x_3)^t**, **c=(c_1,c_2)^t**
    then we can write it just as
        Ax **= c**

    (meanwhile the augmented matrix is "**[A|c]**")

    // give example with numbers

    Now if **A** is an **m x n** matrix, **B** is an **n x k** matrix, and **c** is a **k x 1** column vector, then
        A(Bc)
    makes sense and is some **m x 1** column vector.

    We define the product AB in such a way that
        (AB)c **= A(Bc)**
    for any **c**:

    // There's only one way to do that:
    Definition:
        If **A** is an **m x n** matrix and **B** is an **n x k** matrix, then the _matrix product_ is the **m x k** matrix AB with entries:
            (**AB)_ij = A_i1B_1j + A_i2B_2j + ... + A_inB_nj**

// to see this, consider canonical basis vectors for **c**

Note that this agrees with the definition of Ab above.

We do then generally have

(AB)C **= A(BC)**

(proof omitted)

Example:

Suppose we have the following two systems of linear equations:

```
  x_1 + 2 x_2 - x_3 = y_1
2 x_1 + 3 x_2       = y_2

  y_1 +   y_2 = 3
3 y_1 - 2 y_2 = 6
```

We want to solve for **x_1,x_2,x_3**, i.e. find the values for which the second set of equations hold when we set **y_1** and **y_2** to the values given by the first set of equations.

```
So let A = (1  2  -1)
           (2  3   0)

       B = (1  1)
           (3 -2)

       c = (3)
           (6)
```

We can write the systems in matrix notation as:

Ax **= y**

By **= c**

So really we're trying to solve

B(Ax) = c

which is the same as

(BA)x **= c**

. So if we calculate the matrix product (BA), a 2 **x 2** matrix, we're left with a single system of linear equations which we can solve by Gaussian elimination (the augmented matrix is the 2 **x 3** matrix **[(BA) | c]**).

Definition:

The _transpose_ of an **m x n A** is the **n x m** matrix **A^T** with entries

(**A^T)_ij = A_ji.**

The _trace_ of an **n x n** matrix **A** is **tr(A) := a_11 + a_22 + ... + a_nn.**

Lecture 5: Matrix algebra

// mention "Gauss-Jordan"

Theorem: If **A**, **B**, **C** are matrices, then

(a) **A+B = B+A**
(**b**) (**A+B)+C = A+(B+C)**
(**c**) **A(B+C) = AB+AC**
(**d**) (**A+B**)C **= AC+BC**
(**e**) **A(BC) = (AB)C**

whenever the sizes of **A**, **B** and **C** are such that the additions and multiplications are defined.

Remarks:

These are familiar from properties of normal addition and multiplication. But note that we haven't claimed

AB=BA

which is false in general!

```
e.g. A = (0 1)  B = (0 0)
         (0 0)      (1 0)
```

Similarly, we don't have cancellation
    AB=AC => B=C
for example, A,B as above and $C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

By (b) and (e), we can omit brackets and just write
    A+B+C
and
    ABC
and so on.

Proofs:
    (a) and (b): immediate from corresponding properties of addition of numbers.

    (c): Say A is m x n and B and C are n x k. Then
        (A(B+C))_ij = A_i1 (B_1j + C_j1) + ... + A_in (B_nj + C_nj)
                    = (A_i1 B_1j + A_i1 C_j1) + ... + (A_in B_nj + A_in C_nj)
                    = (AB)_ij + (BC)_ij

    (d): Similar.

    (e): [see insert]

Definitions:
    For any m and n, the _zero matrix_ of size m x n is the matrix 0_mxn (or just 0) with
        (0_mxn)_ij = 0 for all i,j.

    For any n, the _identity matrix_ of size n x n is the matrix I_n (or just I) with
        (I_n)_ij = { 1 if i=j
                   { 0 else

Lemma: For any matrix A
    (a) A+0 = A = 0+A
    (b) A0 = 0
    (c) 0A = 0
    (d) AI = A
    (e) IA = A

    whenever the sizes are such that the operations are defined.
Proof:
    Easy, omitted.

    // do a 2x3 * 3x3 example of (d)

Definition:
    Let A be a square matrix.

    B is an _inverse_ of A if AB = I = BA.

    If such an inverse exists, A is _invertible_.

    If A is not invertible, it is _singular_.

Fact:
    If A is invertible, it has a *unique* inverse.

    The inverse is denoted by A^-1.

/*
Fact:
    Actually, we'll see that if A is square and there exists B such that AB = I, then A is invertible and B=A^-1. Similarly for BA = I.
*/

Fact:
    $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible iff the "determinant" det(A) = ad-bc != 0, and

then

A^{-1} = 1/(ad-bc) (d -b)
(-c a)

// Remark:
// We will define a determinant for **n x n** matrices, and generally a square
// matrix will be singular iff it has det 0

Exercise: check that indeed **AA^{-1} =** I **= A^{-1}A**

Lemma: If **A** and **B** are invertible, then AB is invertible and
(**AB)^-1 = B^-1A^-1**
Proof:
(**B^-1A^-1)(AB) = B^-1(A^-1A)B = B^-1IB = B^-1B=I**
and
(**AB)(B^-1A^-1) =** I
by a similar argument.

Remark: If we have a linear system
Ax=c
and if **A** is invertible, then
x = A^{-1}c
is the only solution. Indeed:

If **x=A^{-1}c**, then by left-multiplying by **A** on both sides of the equation we obtain **Ax=AA^{-1}c**, i.e. **Ax=Ic**, i.e. **Ax=c**. So **x=A^{-1}c** is a solution.

If **Ax=c**, then by left-multiplying by **A^{-1}** on both sides of the equation we obtain **x = A^{-1}c**, so there are no other solutions.

Lecture 6: Elementary matrices

/*
Theorem: Let **r** be a row operation. Then
r(AB) = r(A)B
Proof:
Let **A_i** be the ith row of **A.** Then **r(A)_i** is some linear combination of the **A_i**, say **r_1A_1 + ... + r_nA_n**.

```
Then r(AB) = \Sigma_i r_i (AB)_i
           = \Sigma_i r_i \Sigma_j A_ij B_j
           = \Sigma_i \Sigma_j r_i A_ij B_j
           = \Sigma_j \Sigma_i r_i A_ij B_j
           = \Sigma_j  r(A)_ij B_j
           = r(A)B
```

But I guess that's setting the brow too high.
***/**

Fix a number n; we will work in this section with matrices with **n** rows.

Definition:
If **r** is an elementary row operation, the corresponding _elementary matrix_ **E_r** is **r(I)**, the result of performing **r** on the identity matrix.

Theorem:
Let **r** be an elementary row operation. Let **E_r := r(I).** Then

(**i**) For any **A**, **r(A) = E_rA**
(ii) **E_r** is invertible, and its inverse is an elementary matrix.
Proof:
(**i**) Proof omitted **//** but give examples

(ii) First note that there is an elementary row operation **r^-1** such that
r(r^-1(A)) = A = r^-1(r(A))
for any **A.** Indeed:
(I) if **r** is multiplying a row through by **c != 0**, let **r^-1** be multiplying the same row by **1/c.**

(II) if **r** is swapping two rows, let **r^-1 = r**
(III) if **r** is adding **c** times row **i** to row **j**, let **r^-1** be adding -c times row **i** to row **j**

Now by (**i**),
E_{r^-1}E_r = r^-1(E_r) r^-1(r(I)) = I **=**
r(r^-1(I)) = r(E_{r^-1}) = E_rE_{r^-1}

Corollary:
An nxn matrix **A** is invertible if and only if some sequence **r_1,...,r_k** of elementary row operations reduces **A** to I, i.e. **r_k(...r_2(r_1(A))...)=I**. In this case,
A^{-1} = E_{r_k}...E_{r_1} = r_k(...r_2(r_1(I))...) .

Proof:
By Gauss-Jordan, there are **r_1,...,r_k** such that **R = r_k(...r_2(r_1(A))**...) is in rref. Let **E = E_{r_k}...E_{r_1}**. Then **R =** EA.

E is invertible since each **E_{r_i}** is. So if **A** is invertible, then so is **R=EA**, and if **R** is invertible then so is **A=E^-1R**.

R is rref, so **R** is invertible if and only if **R=I**: indeed, if **R !=** I, then the last row of **R** is zero, but then the last row of RC is zero for any **C**, so **R** is not invertible.

So **A** is invertible if and only if EA **=** I, and then:
EA **=** I **=> EAA^-1 = IA^-1 => E = A^-1**
so **E** is the inverse of **A**.

Using this to calculate inverses:
Given a square matrix **A**, apply Gauss-Jordan to get an rref, and simultaneously perform the row operations on I. **A** is invertible iff the rref we get is I, and in that case we will have transformed I to **A^{-1}**.

```
e.g. ( 1 2 3 )  ( 1 0 0 )
     ( 4 5 6 )  ( 0 1 0 )
     ( 7 8 9 )  ( 0 0 1 )
```

Lecture 7

The nicest kind of behaviour for a linear system
Ax=b
is when there _exists_ a _unique_ solution, for any **b**.

We saw that we have this if **A** is invertible.

There are two ways it could fail: there could be **b** for which we have no solutions, or there could be **b** for which we have more than one solution.

But for square **A (i.e. n equations in n unknowns)**, if **A** is not invertible then they ***both*** fail:

Do example with
(1 2 1)
(2 1 2)
(1 1 1)

A has rref **R=EA** with **E** invertible, and if **A** is not invertible then **R !=** I.

But then some column of **R** doesn't have a leading 1, so we get infinitely many solutions to **Rx=0** (the variable corresponding to that column can take ***any*** value), and for each such **x** we have **Ax=E^-1Rx=E^-10=0**,
so **Ax=0** has infinitely many solutions.

Also the last row of **R** is zero, so the last entry of Rx is zero for any **x**, so if the last entry of **b** is not zero then **Rx=b** has no solutions. But then **Ax=E^-1b** has no solutions, since if **Ax=E^-1b** then **Rx=EAx=EE^-1b=b**.

Theorem:
Let **A** be an a square matrix. Then the following are equivalent:
(a) **A** is invertible

(**b**) **A** has rref I
(**c**) **A** is a product of elementary matrices
(**d**) **Ax=b** has a solution for any **b**
(**e**) The only solution to **Ax=0** is **x=0**
(**f**) **Ax=b** has a unique solution for any **b**

Proof:

(a) **<=> (b)**:
Proved above.

(**c**) **=> (a)**:
Elementary matrices are invertible, so so are products of them.

(a) **=> (c)**:
If **A** is invertible, we saw above that **A^{-1}=E_{r_k}...E_{r_1}.**
So **A=E_{r_1}^{-1}...E_{r_k}^{-1}**, which is a product of elementary matrices.

(a) **=> (f)**:
We saw this earlier.

(**f**) **=> (d)**:
Blatant.

(**f**) **=> (e)**:
Clear.

(**e**) **=> (a)**:
Let **R =** EA be rref, **E** a product of elementary matrices. **E** is invertible, so **Ax=0** iff **EAx=E0** iff **Rx=0.** If **R!=I**, then there are infinitely many solutions to **Rx=0**; indeed, we get a solution for any choice of values of the variables corresponding to columns which don't contain a leading 1.

(**f**) **=> (a)**:
Let **R =** EA as in the previous implication. If **R!=I**, then some row of **R** is 0. But then the corresponding row of Rx is 0 for any **x**, so if **b** is not zero in that row then **Rx=b** has no solutions, so **Ax=E^{-1}b** has no solutions.

Remark:

If **A** isn't square, this doesn't work

Example:

Suppose **F : R^2 -> R^2** is differentiable.

Then at a point **a=(a_1,a_2)**, there is some direction such that if **z** moves from a in that direction then **F(z)** moves from **F(a)** vertically upward (instantaneously) unless there's a direction such that if **x** moves from a in that direction then **F(z)** remains constant (instantaneously).

Proof: consider "Jacobian" of **F** at a:

Jac(F)_a = (dF_1/dx dF_2/dx) where **F(x,y) = (F_1(x,y),F_2(x,y))**
(**dF_1/dy dF_2/dy**)

then if **v** is a column vector representing a velocity, then **Jac(F)v** is the velocity of **F(z)** as **z** moves from a with velocity **v**

Summary of what we've done so far

Given a linear system **Ax=b**:

Gauss-Jordan yields a sequence of elementary row operations which when applied to **A**, give an rref matrix **R.**

Applying those row operations to I yields an invertible matrix **E** such that
EA=R .
E is a product of the elementary matrices corresponding to the elementary row operations, and **E^-1** is the product of the elementary matrices corresponding to the inverse row operations.

Now because **E** is invertible,
Ax=b <=> EAx = Eb **<=> Rx=Eb**

Augmented matrices are just a way of simultaneously calculating **R** and Eb.

A is invertible if and only if **R=I**, and then since **EA=R=I** we have **E=A^-1.**

So then
Rx=Eb

is saying
$x=A^{-1}b$

If **R** has a row of zeros, then
$Rx=Eb$
isn't always solvable, so nor is
$Ax=b$.

If **R** has a column which doesn't contain a leading 1, then
$Rx=0$
has multiple solutions, so also
$Ax=0$
has multiple solutions.

If **A** is square but not invertible, then **R** has a row of zeros and a column which doesn't contain a leading 1.

Diagonal, triangular and symmetric matrices

Definition:
A _diagonal_ matrix is a square matrix **D** with $D_{ij}=0$ if $i \neq j$.

$Diag(a_1,...,a_n)$ is the nxn diagonal matrix **D** with $D_{ii}=a_i$.

An _upper triangular_ matrix is a square matrix **T** with $T_{ij}=0$ if $i>j$.

A _lower triangular_ matrix is a square matrix **T** with $T_{ij}=0$ if $i<j$.

Remarks:
$Diag(a_1,...,a_n)Diag(b_1,...,b_n) = Diag(a_1b_1,...,a_nb_n)$.
In particular, diagonal matrices commute with each other.

The product of two upper triangular matrices is also upper triangular.

Similarly for lower triangular.

The result of Stage I of Gauss-Jordan elimination (a "row echelon form" matrix) is upper triangular.

Definition:
A _symmetric_ matrix is a square matrix **A** such that $A^T=A$.

Recall:
$(A^T)_{ij} = A_{ji}$

Lemma:
(i) For any matrix **A**, $(A^T)^T = A$
(ii) For any matrices **A** and **B** such that AB is defined,
$(AB)^T = B^TA^T$
// demonstrate with an example, using colours

(iii) If **A** is invertible, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.
(indeed: $(A^{-1})^T A^T = AA^{-1} = I = A^{-1}A = A^T (A^{-1})^T$)

Remarks:
If **A** and **B** are symmetric, then so is $A+B$.

If also $AB=BA$, then AB is symmetric, since $(AB)^T = B^TA^T = BA = AB$.

Conversely, if AB is symmetric then $AB=BA$, since
$BA=B^TA^T=(AB)^T=AB$.

If **A** is symmetric and invertible, then $(A^{-1})^T = (A^T)^{-1} = A^{-1}$, so A^{-1} is symmetric.

For any **A**, AA^T is symmetric, since $(AA^T)^T = (A^T)^TA^T=AA^T$.
Similarly A^TA is symmetric.

Determinants

Call a selection of cells of a square matrix "acceptable" if each row and column contains exactly one selected cell.

An acceptable selection has a "sign", positive or negative. The principal diagonal (top-left to bottom-right) is positive. Swapping two rows or columns of an acceptable selection produces an acceptable selection of the opposite sign.
(It isn't obvious that this notion makes sense, but it does!)

The "signed product" of a acceptable selection is the product of the entries if the selection is positive, or minus the product of the entries if the selection is negative.

The _determinant_ **det(A)** of a square matrix **A** is the sum of the signed products of all the acceptable selections.

Remarks:

det(I) = 1, and generally **det(Diag(a_1,a_2,...,a_n)) = a_1a_2...a_n**, since every selection other than the principal diagonal has a 0 entry.

If **A** has a zero row or a zero column, then every selection has a 0 entry so **det(A)=0**.

Transposing a selection doesn't change its sign, so
det(A) = det(A^T)

If we swap two rows or columns, the selections flip sign, so the determinant is multiplied by -1.

In particular, if **A** has two equal rows or columns, then **det(A)=-det(A)**, so **det(A)=0**.

If we consider varying one row or column while not keeping the rest of the matrix fixed, the determinant acts "linearly", meaning

$$\det\begin{pmatrix} ca11 & ca12 & ca13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix} = c \det\begin{pmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix}$$

for any number **c**, and

$$\det\begin{pmatrix} a11+b11 & a12+b12 & a13+b13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix} =$$

$$\det\begin{pmatrix} a11 & a12 & a13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix} + \det\begin{pmatrix} b11 & b12 & b13 \\ a21 & a22 & a23 \\ a31 & a32 & a33 \end{pmatrix}$$

(and similarly for any size of square matrix and for any choice of **row/column**)

This is easy to see by thinking about the selections, e.g. we have
(a11+b11)a22a33 = a11a22a33 + b11a22a33

In particular, **det(cA) = c^nA** for an nxn matrix **A**, because we're multiplying each of the **n** rows of **A** by **c** to get to cA.

Note that **det(A+B)** is generally ***not*** equal to **det(A)+det(B)**
(and there's no nice formula).

So we see we have the following behaviour under elementary row operations:

If **B** is the result of multiplying a row of **A** by **c!=0**, then
det(B) = c det(A)
If **B** is the result of swapping two rows of **A**, then
det(B) = -det(A)
If **B** is the result of adding a multiple of a row of **A** to another row of **A**,
det(B) = det(A)

In each case, the determinant is multiplied by a non-zero number (**c**, -1, or 1).

Calculating determinants by row reduction:

Reduce **A** to an rref **R** by **e**.r.o's. Then **det(R)=e det(A)** where **e!=0** is the product of the numbers the **e**.r.o's multiply the determinant by (**c** or -1 or 1, as above). If **R=I**, then **det(R)=1**, and then **det(A) = 1/e**. Otherwise, **R** has a row of zeros so **det(R)=0**, so **det(A) = 0/e = 0**.

Example:
```
 A= (4 3)  ->  (1 3/4)  ->  (1  3/4)  ->  (1 3/4)  ->  (1 0)
    (2 1)      (2  1 )      (0 -1/2)      (0  1 )      (0 1)
        *(1/4)                     *(-2)
so (-2)(1/4)det(A)=det(I)=1
so det(A) = -2
```

Example:
```
    A = (0 1)  ->  (1 0)
        (1 0)      (0 1)
              *(-1)
    so det(A) = -1
```

Since **R=I** iff **A** is non-singular, we can deduce

Theorem: For any square matrix **A**,
det(A)=0 if and only if **A** is singular.

Also

Theorem: If **A** and **B** are nxn matrices, then
det(AB) = det(A)det(B)

Proof:
If **A** is singular, then there is a **b** such that **Ax=b** has no solution, so nor does **ABx=b**, so AB is singular. So the formula holds.

If **A** is non-singular, then **A=E** a product of elementary matrices. So **A** is the result of applying the corresponding row operations to I, and AB is the result of applying them to **B**. Since each row operation multiplies the determinant by a number depending only on the row operation, **det(AB)=det(A)det(B).**

Cofactor expansion:
// demonstrate with a 4x4 matrix: consider acceptable selections when we
// fix 1,1 to be selected, note they are the acceptable selections of the
// minor; sim for 1,2, but note that the signs get flipped, and so on

Let **A** be an nxn matrix.

Cofactor expansion on the first row:
Let **M_ij** be the (**n-1)x(n-1)** matrix you get if you cross out the ith row and the the jth column of **A**. Then
det(A) = a_11 det(M_11) - ... **+ a_1n (-1)^(1+n) det(M_1n)**
a_11 (-1)^(1+1) det(M_11) + ... + a_1n (-1)^(1+n) det(M_1n)

This works on an arbitrary row, but the signs get flipped if it's an odd row, so if we expand through the ith row we get:
det(A) = a_i1 (-1)^(i+1) det(M_i1) + ... + a_in (-1)^(i+n) det(M_in)

Let **C_ij = (-1)^(i+j) M_ij**, the **ij^th _cofactor_.** Then
det(A) = a_i1 C_i1 + ... + a_in C_in .

Note that if we use the cofactors of a ***different*** row, we get zero:
a_i1 C_j1 + ... + a_in C_jn = 0 if **i != j**
since that's the determinant of the matrix you get if you change the jth row to equal the ith row.

Similarly we can expand through a column:
det(A) = a_1j C_1j + ... + a_nj C_nj .
(and again we get zero if we use a different column)

Now if we let **adj(A)** be the matrix with entries
adj(A)_ij = C_ji
we can write these as matrix multiplications:
A adj(A) = det(A) I **= adj(A) A**

hence

Theorem: if **A** is invertible, then
A^-1 = (1/det(A)) adj(A)

Eigenfoo

* Matrices as linear transformations
* Diagonal matrices as giving nice description of corresponding transformation
* Change of basis and "secretly diagonal" matrices.

* Hmm, but I don't see how to explain it understandably in those terms without effectively going through the chapters we're skipping...
* Maybe try to keep it elementary by talking about "linear changes of coordinates"?

Consider the **x**-y plane, and consider reflection in the **x**-axis as a map from the plane to itself.

So the point **(x,y)** is mapped to **(x,-y)**.

Representing the point **(x,y)** by the column vector **(x,y)^T**, we can represent the map by the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Now consider reflection about the line **y=2x.**

(1,0) is mapped to **(-3/5, 4/5**)
(0,1) is mapped to **(4/5,3/5)**

so this has matrix representation $\begin{pmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{pmatrix}$

but if we pick new co-ordinates, such that the column vector **(1,0)^T** now represents the point with **x**-y co-ordinates **(1,2)**, which is on the line, and **(0,1)^T** represents the point with **x**-y co-ordinates **(-2,1**), which is perpendicular to the line, so generally **(u,v)^T** represents the point with **x**-y co-ordinates **(u-2v,2u+v)**, then **(u,v)^T** is mapped to **(u,-v)^T**, so the matrix representation with respect to this new co-ordinatisation is again $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

More generally: a map represented by a diagonal matrix is particularly simple to understand - it "rescales" the co-ordinates.

$$\begin{pmatrix} l1 & 0 & 0 \\ 0 & l2 & 0 \\ 0 & 0 & l3 \end{pmatrix} \begin{pmatrix} d \\ e \\ f \end{pmatrix} = \begin{pmatrix} l1\ d \\ l2\ e \\ l3\ f \end{pmatrix}$$

A matrix which isn't diagonal may nonetheless represent a map which is diagonal if we pick the co-ordinatisation cleverly. Such a matrix is called "diagonalisable".

// it will turn out, after passing to complex numbers, that ***every*** matrix
// is ***almost*** diagonalisable

Eigenvectors

A column vector **x** is an _eigenvector_ of a matrix **A** if **x!=0** and there exists a number **l** such that

Ax **=** lx ;

l is then called the _eignenvalue_ of **x**, and **x** is an _l-eigenvector_ of **A**, and **l** is an _eigenvalue of **A**_.

Examples:

(0,1,0)^T is a 2-eigenvector of **A=Diag(1,2,3)**, and **(0,0,-5)^T** is an eigenvector with eigenvalue 3. **(1,1,0)^T** is not an eigenvector.

We saw above that **(1,2)^T** is a 1-eigenvector of **A=** $\begin{pmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{pmatrix}$

Each non-zero vector on the line **y=2x** is an eigenvector with eigenvalue 1, and each non-zero vector on the perpendicular line **y=-x/2** is an eigenvector with eigenvalue -1. There are no other eigenvectors.

A reflection in a plane through the origin in three−space will have every vector on the plane as a 1−eigenvector, and every vector on the normal line as a −1−eigenvector, and no other eigenvectors.

A rotation in three−space about an axis through the origin will have every non−zero vector on the axis as a 1−eigenvector, and (unless the angle is a multiple of **\pi**) no other (real) eigenvectors.

Any **x!=0** is a 0−eigenvector of the zero matrix.

Any **x!=0** is a 1−eigenvector of the identity matrix.

// Idea: if **A** is "secretly diagonal", the eigenvectors will show us how to
// find an appropriate co−ordinatisation.

Remarks: if **x** is a **\lambda**−eigenvector of **A**, and **c!=0** is a number, then cx is also a **\lambda**−eigenvector. If **y** is another **\lambda**−eigenvector, then **x+y** is a **\lambda**−eigenvector (assuming **x+y!=0**).

// If **y** is a **\mu**−eigenvector and **\mu != \lambda**, then **x+y** is ***not*** an
// eigenvector!

Definition:
The _l−eigenspace_ of **A** is the set of **l**−eigenvectors,
i.e. the set of solutions of the equation
Ax **=** lx
i.e. the set of solutions of the equation
(**A**−lI)x **= 0**

So **l** is an eigenvalue iff the **l**−eigenspace contains a non−zero point.

Finding eigenvalues and eigenvectors

\lambda is an eigenvalue of an nxn matrix **A**
iff there exists **x!=0** such that **Ax=lx**
iff there exists **x!=0** such that (**A**−lI)x **= 0**
iff the matrix (**A**−lI) is singular
iff **det(A−lI)=0.**

So the eigenvalues of **A** are precisely the zeros of the function of **l**
\xi_A(l)=det(A−lI)

\xi_A is a polynomial of degree **n** − it looks like
\xi_A(l) = l^n+c_{n−1}l^{n−1}+...+c_1l + c_0.
It is called the _characteristic polynomial_ of **A.**

```
// 2x2 example
// example with A = (0 0  0)
//                  (1 2 −1)    \xi_A(l) = l((2−l)(6−l)+4) = l^3 − 8l^2 + 16l =
//                  (1 4  6)                l(l−4)^2
```

Once we've found the eigenvalues **l_1,...,l_s**, we can find the eigenspaces by solving the equations
(**A−l_iI)x=0.**

Continuing the above example:
the eigenvalues are 0 and 4

The 0−eigenspace is the set of solutions to
(**A−0I)x=0**
i.e. to
Ax=0

Solve by Gaussian elimination:
[flip 0s to bottom; subtract 1st from 2nd to get (0 −2 −7); so get

```
(1 4  6 )
(0 −2 −7)
(0 0  0 )
```

$$\begin{pmatrix} 1 & 4 & 6 \\ 0 & 1 & \mathbf{7/2} \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & -8 \\ 0 & 1 & \mathbf{7/2} \\ 0 & 0 & 0 \end{pmatrix}$$ solution set: **(8t,−7t/2,t)**

The 4−eigenspace is the set of solutions to

$$\begin{pmatrix} \mathbf{-4} & \mathbf{0} & \mathbf{0} \\ 1 & -2 & -1 \\ 1 & 4 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \mathbf{1/2} \\ 0 & 4 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \mathbf{1/2} \\ 0 & 0 & 0 \end{pmatrix}$$ solution set: **(0,−t/2,t)**

Another example, where we get a higher−dimensional eigenspace:

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

\xi_A = (2−l)^3

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

(t,0,s)

Diagonalisation

Definition:
nxn matrices **A** and **B** are _similar_ if there exists an invertible nxn matrix **P** such that **A = P^−1BP**.

An nxn matrix **A** is _diagonalisable_ if it is similar to a diagonal matrix, i.e. there is an invertible matrix **P** such that **P^−1AP** is diagonal.
We then say that **P _diagonalises_ A**.

What this really means (not on syllabus):
A and **B** are similar iff they are representations of the same map of **n**−space to itself, with respect to possibly different co−ordinatisations.

P is the "change of co−ordinates" matrix, which maps the co−ordinates of a point according to the co−ordinatisation **B** uses to the co−ordinates of that same point according to the co−ordinatisation **A** uses.

Remarks:
If **A** and **B** are similar, say **A=P^−1BP**, then
(**i**) **det(A)=det(B)**
since **det(A)=det(P^−1BP)=det(P^−1)det(B)det(P)=det(P)^−1det(B)det(P)=det(B)**
(ii) **A** and **B** have the same eigenvalues
since if **Ax=lx** and **x!=0**, then **P^−1BPx=lx** so **BPx=Plx=lPx**, and **Px!=0** since **P** is invertible.

Example of why this is useful:
If **P^−1AP = D**, so **A = PDP^−1**, then **A^2 = PDP^−1PDP^−1 = PD^2P^−1**, and generally **A^k = PD^kP^−1**. Working out **D^k** is easy − just raise each diagonal entry to the kth power.

So if **A** is diagonalisable, and we can work out the **P** and **D** involved, then we can (quite) easily calculate powers of **A**.

c.f. discrete dynamical systems (next week)

How to diagonalise:

Given an nxn matrix A, try to find n eigenvectors $p_1,\dots,p_n$ of A such that the matrix $P = [p_1 \dots p_n]$, whose columns are those eigenvectors, is invertible.

If you can do this, then A is diagonalisable, and P diagonalises it, and the diagonal entries of $D=P^{-1}AP$ are the eigenvalues of the p_i.

Indeed, say p_i is a l_i-eigenvector; then $P^{-1}APe_1 = P^{-1}Ap_1=P^{-1}lp_1=lP^{-1}p_1=le_1$, so the first column of $P^{-1}AP$ is le_1. Similarly for the other columns - the ith column has l_i in the ith row and is zero elsewhere.

If you can't do this, then A isn't diagonalisable - because if A were diagonalisable, say $P^{-1}AP=D$, then $A=PDP^{-1}$ and it's not hard to see that the columns of P would be eigenvectors.

Example (do simultaneously with theory?):

$A = \begin{pmatrix}1 & 2\\ 5 & 4\end{pmatrix}$ $\quad \xi_A(l) = (l-1)(l-4) - 10 = l^2-5l-6 = (l-6)(l+1)$

Solve $(6I-A)x=0$: $x = (2t/5,t)^T$

Solve $(-1I-A)x=0$: $x = (-t,t)^T$

Take $p_1 = (2,5)$ and $p_2 = (1,-1)$, say.

Then $P = \begin{pmatrix}2 & 1\\ 5 & -1\end{pmatrix}$ is invertible since $\det(P) = -2-5 \neq 0$

$P^{-1}AP(1,0)^T = P^{-1}A(2,5)^T = P^{-1}\, 6(2,5)^T = 6\, P^{-1}(2,5)^T = 6\,(1,0)^T = (6,0)^T$

$P^{-1}AP(0,1)^T = (0,-1)^T$

$\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix} = \begin{pmatrix}a\\ c\end{pmatrix}$

So $D := P^{-1}AP = \begin{pmatrix}6 & 0\\ 0 & -1\end{pmatrix}$

Example of undiagonalisable A:

$A = \begin{pmatrix}-1 & 1\\ -1 & 1\end{pmatrix}$ $\quad \xi_A(l) = (l+1)(l-1)+1 = l^2 \quad Ax=0 \iff x=(t,t)^T$

So a matrix of eigenvectors would look like

$P = \begin{pmatrix}t & s\\ t & s\end{pmatrix}$

which is not invertible, since the columns are multiples of each other.

Finding an invertible matrix of eigenvectors:

If $P=[p_1 \dots p_n]$ is to be invertible, we need at least that no two of the eigenvectors we pick are multiples of each other ($p_i=cp_j$). In simple examples, that will be enough.

More generally, we'll see later that P is invertible iff $p_1,\dots,p_n$ "span" R^n, meaning that ***any*** column vector x can be written as a sum of multiples of the p_i:

$x = c_1p_1 + \dots + c_np_n$

So we'll see that A is diagonalisable iff any vector is a sum of eigenvectors.

Remark:

If A is diagonalisable, say $D=P^{-1}AP$ is diagonal, then since D and P are similar they have the same eigenvalues. So the diagonal entries of D are precisely the eigenvalues of A.

So P has to have an eigenvector of each eigenvalue.

(It might have more than one of a given eigenvalue)

Fact:

If the p_i have different eigenvalues, then $P=[p_1,\dots,p_n]$ is invertible.
// We'll understand why later.

So

Theorem:

If an nxn matrix A has n distinct eigenvalues,

i.e. if the characteristic polynomial **\xi_A(l)** factors into distinct linear factors,
then **A** is diagonalisable.

Example:

```
(1 1 0)
(0 2 0) \xi_A(l) = (l-1)(l-2)(l-3) so the eigenvalues are 1,2,3
(0 0 3)
```

3 distinct eigenvalues, so diagonalisable.

Linear discrete dynamical systems

Example - population dynamics:

The caves on mars contain intriguing vegetation and three species of fauna: the Squongles, the Squoogles and the Squeegles.

Left alone, each species multiplies rapidly, increasing 30% per Martian year.

However, the Squeegles sometimes consume the Squoogles, who hunt the Squongles, who are peaceful but tend to accidentally step on the Squeegles.

A scientific mission is planned to investigate one of the caves. Its current population is estimated to be 0.8m Squongles, 0.6m Squoogles, 0.4m Squeegles.

In this cave, on average a Squeegle kills off a Squoogle once every two years. So the Squoogle population is reduced each year by 50% of the Squeegle population. Similarly for Squoogles killing Squongles and Squongles killing Squeegles.

You are evil, and work for the bioweapons division of your government, which does not want the existence of life on mars revealed to the public. Killing off large numbers of Martians is difficult, but you could introduce further Squongles, Squoogles and Squeegles from other caves.

Using your eigen-fu, can you find a way to reduce the population sufficiently that astro-agents with pointy sticks will suffice to finish off the rest?

The Squongles have started deliberately trying to stamp on Squeegles, and now each kill 1 Squeegle per year. How does this change things?

//(real evalue: 0.57; evector: (0.44, 0.56, 0.70)^T)

What happens if we introduce a 4th species, and change the matrix to

```
(1.5 -0.5   0      0)
(0    1.5   -0.5   0)
(0    0     1.5    -0.5)
(-0.5 0     0      1.5)
```

real epairs: **(1,(1,1,1,1)**); **(2,(1,-1,1,-1)**)

Remark: these matrices are not diagonalisable over the reals (although they are over the complex numbers! See later)

Example - network analysis and centrality:

Suppose we have a network consisting of some nodes and links from nodes to other nodes.

e.g. webpages and weblinks
journals and citations
"tweeters" and "following"

We want to determine which nodes are most "important" in the network.

Idea: an important node is one which is linked to from important nodes.

This seems circular! It is, but it can make sense anyway...

Say we have five webpages, with the following link graph:

Take a large number of people. Put each in front of a computer with a browser loaded to a random one of our five pages. Suppose that they're really just browsing aimlessly, and they click links at random.

Question: where will they end up?

Definition:

A column vector is _stochastic_ if each entry is non-negative and the sum of the entries is 1.

Idea: we can interpret a stochastic vector as giving the probability of being in each of **N** states, when we know we have to be in one of them.

A matrix is _stochastic_ if all its columns are.

Idea: the jth column represents the probabilities of something currently in the jth state changing to each of the states in the next step.

Fact [Perron-Frobenius for stochastic matrices]:
If **A** is a stochastic square matrix and for some **n**, all entries of **A^n** are positive, then it has a unique stochastic eigenvector **v_l**, which has **e**-value 1, and for any stochastic **x**

lim_{n -> \inf} A^n x = v_l

Stochastic diagonalisable example:

```
( 0    1  2/3)
( 1/3  0  1/3)
( 2/3  0  0  )
```

(e.g. blinking links!)

evalues and evectors are:
1, **-2/3**, **-1/3**
with corresponding evectors
(9, 5, 6) evalue 1

(1, 0, -1) evalue **-2/3**

(1, 1, -2) evalue **-1/3**

So stochastic 1-e-vector is **v_l = (9/20, 5/20, 6/20)^T**

So this is the limit behaviour.

Any stochastic **e**-vector **v** can be written as
v = v_l + x_2 + x_3
where **x_2** is a **(-2/3)**-evector and **x_3** is a **(-1/3)**-evector.

e.g. **(1/3,1/3,1/3) = v_l + (1/12, 1/12, -1/6)^T + (1/5, 0, -1/5)^T**

So **A^n v =** ...

Another:

```
( 1/4 1/3 1/3)      ( 2/3  0  -3/2 )
( 1/4 1/3 1/6)   P= ( 1/2 -1   1/2 )   D=diag(1,1/6,-1/12)
( 1/2 1/3 1/2)      (  1   1    1  )
```

evalues and evectors are:
1, **1/6**, **-1/12**
with corresponding evectors
(**2/3**, **1/2**, 1) evalue 1

(0, -1 , 1) evalue **1/6**

(-3/2, **1/2**, 1) evalue **-1/12**

So stochastic 1-e-vector is (**4/13 3/13 6/13)^T**

Things I should say about this stuff after reading week (maybe folding in to complex eigenvalues stuff):
- define "Markov chain"
- physical example, and connections to differential equations?
- the coolness of applying low-dimensional geometric intuition to high-dimensional wholly non-geometric situations

Complex numbers

Example: simple harmonic motion

A body of mass 1kg is constrained to a line, and connected by a spring to a fixed point we'll call 0. Let **p(t)** be the position at time **t**, and **q(t)** the **momentum=velocity**. Say **p(0)=0**, **q(0)=1**.
//Just draw a diagram for all that.
Suppose the strength of the spring is such that the force acting on the body is −p. So we have

$$dp/dt = q \qquad dq/dt = -p.$$

The "correct" thing to do would be to solve this using the theory of differential equations.

But let's try a discrete approximation. Let Dt **> 0** be quite small, say Dt **= 1/100**.

Write $p_n = p(nDt)$, $q_n = q(nDt)$, $x_n = (p(nDt), q(nDt)$.

Suppose we know x_n. Then what is x_{n+1}?

In Dt seconds, the position will have increased by approximately qDt.
Meanwhile the velocity will have increased by approximately −pDt.

So x_{n+1} is roughly $(p_n + Dt\, q_n, q_n - Dt\, p_n) = A x_n$
where $A = \begin{pmatrix} 1 & Dt \\ -Dt & 1 \end{pmatrix}$

// show it in markovvis, note periodicity

$$\xi_A(l) = (l-1)^2 + (Dt)^2 = l^2 - 2l + Dt^2 + 1$$

quadratic formula: zeroes are

$$\begin{aligned}(-b \pm \sqrt{b^2-4ac})/2a &= (2 \pm \sqrt{(-2)^2 - 4(Dt^2 + 1)})/2 \\ &= 1 \pm \sqrt{-4Dt^2}/2 \\ &= 1 \pm \sqrt{-Dt^2}\end{aligned}$$

so no solutions.

So no **e**−values, no **e**−vectors, our techniques are useless.

Or are they?

Complex numbers

Basic idea: declare that negative numbers ***do*** have square roots, and see what happens.

Introduce a new symbol **i** to be a square root of −1, so we declare $i^2 = -1$.

Now we try to extend the usual real numbers and their addition and multiplication to a well−behaved class of numbers which includes **i**, where 'well−behaved' means that we have all the usual nice properties:

$$0+z = z$$
$$z+w = w+z$$
$$1z = z$$
$$zw = wz$$
$$z(w+u) = zw + zu$$

If **i** is one of these numbers, then so must be bi for **b** real, hence so must be **a+bi** for a and **b** real. We'll see that we can stop there; so we declare the _complex numbers_ to consist of the numbers **a+bi** where a and **b** are real.

$$a+bi = c+di \text{ iff } a=c \text{ and } b=d$$

$Re(a+bi) = a$
$Im(a+bi) = b$
A complex number **z** with $Im(z)=0$ is _**real**_.
A complex number **z** with $Re(z)=0$ is _imaginary_ (e.g. **i**, −5i, but not **1+i**).

Now let's define addition and multiplication of complex numbers such that the laws above hold.

We have to have

$$(a+bi) + (c+di) = a+c + bi+di = (a+c) + (b+d)i$$

and

$$\begin{aligned}(a+bi)(c+di) &= (a+bi)c + (a+bi)di \\ &= ac + bci + adi + bdi^2 \\ &= ac - bd + bci + adi \qquad (i^2=-1) \\ &= (ac-bd) + (ad+bc)i\end{aligned}$$

One can check that the above laws then do hold.

Argand diagram.

Polar form

Remark:
For any complex number $z=a+ib$, iz is the point on the Argand diagram you get by rotating z by $\pi/2$ counterclockwise around the 0.

$$iz = -b + ai$$

Definition: for θ real,

$$e^{i\theta} = \cos\theta + i \sin\theta$$

= [point on unit circle with angle θ counterclockwise from positive real line]

Lemma:

$$e^{i\theta}e^{i\phi} = e^{i(\theta+\phi)}$$

Proof:
If $\phi=0$, $e^{i\phi}=1$ so obvious.
If $\phi=\pi/2$, $e^{i\phi}=i$, so

$$e^{i\theta}e^{i\phi} = e^{i\theta}i = ie^{i\theta}$$

which, as we just saw, is $e^{i\theta}$ rotated by $\pi/2$, which is i rotated by θ.

So multiplication by $e^{i\theta}$ acts as rotation by θ on 1 and i.

So if $e^{i\phi} = a+ib$, then

$$\begin{aligned}e^{i\theta}(a+ib) &= e^{i\theta} a + e^{i\theta} ib \\ &= [a \text{ rotated by } \theta] + [ib \text{ rotated by } \theta] \\ &= [a+ib \text{ rotated by } \theta] \\ &= [e^{i\phi} \text{ rotated by } \theta] \\ &= e^{i(\theta+\phi)}\end{aligned}$$

Alternative proof using trig laws:

$$\begin{aligned}e^{i\theta}e^{i\phi} &= (\cos\theta + i \sin\theta) (\cos\phi + i \sin\phi) \\ &= \cos\theta \cos\phi - \sin\theta \sin\phi + (\sin\theta \cos\phi + \cos\theta \sin\phi) i \\ &= \cos(\theta+\phi) + i \sin(\theta+\phi) \\ &= e^{i(\theta+\phi)}\end{aligned}$$

[
Now define for an arbitrary complex number $a+ib$:

$$e^{a+ib} = e^a e^{ib}$$

Then we have $e^{z+w} = e^z e^w$.

Also note : $\frac{d}{d\theta} e^{i\theta} = ie^{i\theta}$
]

Any non-zero complex number z can be written "in polar form" as $re^{i\theta}$ for some real r and θ. r is the _absolute value_ of z, written $|z|$, and θ is an _argument_ of z, written $\arg(z)$.

By Pythagoras, $|a+ib| = \sqrt{a^2+b^2}$.

arg(a+ib) = arctan(b/a)

Note that the argument is not uniquely defined, since if **n** is an integer then
e^i(\th+2n\pi) = e^i\th (e^2i\pi)^n = e^i\th 1^n = e^i\th .

Multiplication in polar form:
r_1 e^i\th_1 r_2 e^i\th_2 = r_1r_2 e^i(\th_1+\th_2)

Now we can see that every non-zero complex number has a (unique) reciprocal:
(**re^i\th**) (**r^-1e^{-i\th}**) = **1**
so we can divide complex numbers by non-zero complex numbers.

Definition: the _complex conjugate_ of **z = a+ib** is **\bar{z} =** a-ib
In polar form, **\bar{re^i\th} = re^{-i\th}**.

Remark:
z\bar{z} = |z|^2
so **1/z = \bar{z}|z|^-2**
so **w/z = w\bar{z}|z|^-2**

Taking roots:
The solutions to **z^n=1** are **e^(k2\pi i / n) k={0,...,n-1}**.

e.g. **\omega := e^(2\pi i / 3) = 1/2 + \sqrt(3)i/2**
and **\omega^2 = e^(2 2\pi i / 3) = 1/2 - \sqrt(3)i/2**
are cube roots of 1.

z^n=re^i\th has **r^{1/n}e^(i\th/n)** as a solution.

The solutions are **r^{1/n}e^(i\th/n + k2\pi i / n) k={0,...,n-1}**.

so **z^n-\alpha** has **n** solutions for **\alpha != 0**...

Fact - the Fundamental Theorem of Algebra:
A polynomial of degree **d** with complex number coefficients splits into **d** linear factors with complex number coefficients.

In particular, the characteristic polynomial of an nxn matrix always has **n** complex roots (counting multiplicities)...

Complex eigenfoo
================

Example:
(0 1 0)
(0 0 1)
(1 0 0)

\xi_A(l) = l^3 - 1

zeroes: **l = 1**, **l = e^(2\pi i / 3) = \omega**, **l = e^(2 2\pi i / 3) = \omega^2**
(**= e^(-2\pi i / 3) = 1/\omega = \bar{\omega}**)

So **A** should be diagonalisable? What is **P**?

e-space of 1 is **(t,t,t)^T**.

What are the **e**-vectors with **e**-value **\omega**?

They're complex!

Fact: Everything we've done so far in linear algebra goes through if we understand "number" as "complex number" rather than "real number"!

So we can allow entries of vectors and matrices to be complex numbers...

Can use the same old techniques to find the **e**-space of **\omega** - solve
(**\omega** I - **A**)x **= 0**

(om -1 0)   (1   0 -om)

```
(0   om  -1)  ->  ...  ->   (0   1  -om^2)
(-1  0   om)                (0   0   0   )
```

solutions are (\omega^2 t, \omega t, t)^T, t any complex number.

Similarly, the \omega^2 e-vectors are (\omega t, \omega^2 t, t)^T.
Note these are the complex conjugates of the \omega e-vectors.

So take P = blah.

What does this have to do with what A does to real vectors?

Well... A is diagonalisable, so every point of C^3 can be written as a sum of e-vectors. In particular, every point of R^3 can be.

```
[ (\omega t, \omega^2 t, t) + (\omega^2 \tbar, \omega \tbar, \tbar) =
    (\omega t + \bar{\omega t}, \omega^2 t + \bar{\omega^2 t}, t+\tbar)
    so get the plane x+y+z=0
]
```

x = \alpha_1 x_1 + \alpha_om x_om + \alpha_ombar x_ombar

x = \xbar = \alpha_1 x_1 + \bar\alpha_om x_ombar + \bar\alpha_ombar x_om

so we must have \alpha_om = \bar\alpha_ombar

```
  Ax = \alpha_1 x_1 + om \alpha_om x_om + ombar \alpha_ombar x_ombar
A^2x = \alpha_1 x_1 + om^2 \alpha_om x_om + ombar^2 \alpha_ombar x_ombar
A^3x = \alpha_1 x_1 + \alpha_om x_om + \alpha_ombar x_ombar
```

Periodicity!

Also, A^3 = PD^3P^-1 = PIP^-1 = I

Simple Harmonic Motion revisited

```
A = ( 1   Dt)
    (-Dt  1 )
```

x_n = (p_n,q_n)^T position-momentum at time nDt
x_{n+1} = Ax_n

\xi_A(l) = (l-1)^2 + (Dt)^2 = l^2 - 2l + Dt^2 + 1

quadratic formula: zeroes are

```
    (-b +/- \sqrt(b^2-4ac))/2a = (2 +/- \sqrt((-2)^2 - 4(Dt^2 + 1)))/2
                               = 1 +/- \sqrt(-4Dt^2)/2
                               = 1 +/- \sqrt(-Dt^2)
                               = 1 +/- iDt
```

so these are the e-values!

Two distinct e-values, so diagonalisable.

What are the e-vectors with e-value i + iDt?

They're complex!

Fact: Everything we've done so far in linear algebra goes through if we understand "number" as "complex number" rather than "real number"!

So we can allow entries of vectors and matrices to be complex numbers...

Can use the same old techniques to find the e-space of \omega - solve ((1+iDt) I - A)x = 0

```
( iDt   -Dt  )    ->   ( 1 i )
(  Dt   iDt  )         ( 0 0 )
```

so (1+iDt) e-vectors are (t,it), t!=0 complex.

similarly, **(t,−it)** are the (1−iDt)−e−vectors.

No real **e**−vectors, but ***every*** vector is the ***sum*** of **e**−vectors, since **A** is diagonalisable.

Indeed:

```
P = (1  1)
    (i -i)

det(P) = -i - i = -2i
P^-1 = (1/-2i) ( -i  -1 )  =  1/2 ( 1  -i )
               ( -i   1 )         ( 1   i )

So (x,y)^T = PP^1(x,y)^T = P ( (x-iy)/2, (x+iy)/2 )^T
           = (x-iy)/2 (1,i)^T + (x+iy)/2 (1,-i)^T
```

(**x** and **y** real or complex)

and then **A(x,y)^T = (1+iDt) (x−iy)/2 (1,i)^T + (1−iDt) (x+iy)/2 (1,−i)^T**

Now **arg(1+iDt) ~=** Dt (for small Dt)
and **|1+iDt| ~= 1**

so if **n ~= 1/Dt**, then (**1+iDt**) **~= (1−iDt) ~= 1**
and so

```
    A^n(x,y)^T = (1+iDt)^n (x-iy)/2 (1,i)^T + (1-iDt)^n (x+iy)/2 (1,-i)^T
               ~= (x-iy)/2 (1,i)^T + (x+iy)/2 (1,-i)^T
               = (x,y)^T
```

So **x_n ~= x_0**.

x_n = position−momentum after nDt **~= 1** seconds

In fact, **r=|1+iDt|** is slightly more than 1. So if **n=1/Dt**, so (**1+iDt)^n** is real, then (**1+iDt^n**) **= r^n** which is slighly more than 1.
Also (**1−iDt^n)=r^n**

So **x_n = r^n x_0**
and **x_{kn} = r^kn x_0**

So our model goes wrong with exponential growth. If we make Dt smaller, then
r^n = (1+n^−2)^(n/2) −−> 1

Vector geometry

R^n consists of **_points_**, which we may specify by co−ordinates
"**P = (x_1,...,x_n)**" or, if you insist, "**P(x_1,...,x_n)**".

O=(0,0,...,0) −− origin

P−>Q is a **_vector_**. Identify with **row/column** vectors.

||v||

v is a _unit vector_ iff **||v|| = 1**

Distance between **P** and **Q = ||P−>Q||**

u.v

u is _orthogonal_ ("at a right angle") to **v** iff **u.v=0**

Geometry of dot product:
Suppose **||v|| = 1**.

w := u − (**u**.v)v is orthogonal to **v**.

proj_v(u) := (u.v)v is called the "orthogonal projection of **u** to **v**".

Can think of **u**.v as the "distance of **u** from **O** in the direction of **v**".

Have right-angled triangle, and deduce **u**.v **= ||u||cos\theta**.

So generally,
u.v **= ||u|| ||v|| cos\theta**

(and really this is the definition of the angle theta between **u** and **v** in a normed vector space)

Planes in **R^3**:
Equation of a plane:
n.x **= c** with **n** a unit vector (the "unit normal" to the plane)
"points of distance **c** from 0 in the direction of **n**"

Distance of **v** from the plane **n.x=c** is **n**.v − **c**.

(Explain that we are here systematically confusing points and vectors)

Parametric form:
If **x_0**, **x_0 + u** and **x_0 + v** are all on a plane, then so is
x_0 + su **+** tv
for any real **s** and **t**, and every point of the plane is of this form.

Finding **u** and **v** for **n.x=0** corresponds to solving the linear equation **n.x=0**...

Geometry of linear systems:
Let **A** be an **m x n** matrix. Let **a_1,...,a_m** be the rows. Then **Ax=0** iff **a_i**.x **= 0** for all **i** iff **x** is orthogonal to each **a_i**.

Solving **Ax=0 (e.g. using Gaussian elimination)** means finding a parametric form for the solution set.

The solution set of **Ax=b** is the translate of the solution set of **Ax=0** by any one solution to **Ax=b**, i.e. if **Ax_0=b** then **Ax=b <=> A(x−x_0)=0**.

Cross product:
Work in **R^3**.

v x w = (det blah,,)

Fundamental property:
For any **u**, **u.(v x w) = det(u v w)**

Hence:
If **v** and **w** are collinear, then **v x w = 0**.
v.(v x w) = 0 = w.(v x w), so **v x w** is perpendicular to **v** and to **w**.

Fact:
det(u v w) = signedVol(parallelepiped of u,v,w)
where the signedVol is positive iff **u**,**v**,**w** is right-hand-oriented
(u thumb; v index; w middle. i,j,k in right-hand oriented)
Proof:
True for transpose of rref - identity matrix corresponds to **i**,**j**,**k**
and any other has a zero column and hence 0 det and 0 volume.

Elementary column operations act on the right like they do on the left.
(sign flips, rescalings, volume-preserving shears)

Suppose **v** and **w** not collinear. Let **^u** be the unique unit vector which is perpendicular to **v** and **w** such that **(^u,v,w)** is right-hand-oriented.

Then **^u** is collinear with **v x w**, so
v x w = (^u.(v x w)) ^u = det(^u v w) ^u
= signedVol(parallelepiped of ^u,v,w) ^u
= area(parallelogram of v,w) ^u
= (||v|| ||w|| sin \theta) ^u
where **\theta** is angle between **v** and **w**

i x j = k

av **x w = a(v x w)**
v x w = -(w x v)
u x (v x w) != (u x v) x w
e.g. **i x (i x j) = i x k =** -j **!= 0 = (i x i) x j**

Example: torque
If a force **F** acts on a body at position **p**, it results in a rate of change of angular momentum **L** of the body around 0 of
dL/dt = p x F
here **L** is a vector giving the axis through 0 around which the body rotates right-handedly, with the rate of rotation proportional to the length of the vector.

Example: Lorentz force
If a particle with charge **q** moves at velocity **v** in a magnetic field **B**, it experiences a force
F = q (v x B)
so if **B** is constant, and there are no other forces, the particle traces out a helix.

(Fact: for any square **A**, **det(A)** is the scaling factor on (signed) volume:
signedVol(Av_1,...,Av_n) = det(A) signedVol(v_1,...,v_n)
(and for any measurable subset **M** of **R^n**, **vol(A(M)**) **= |det(A)| vol(M)**))

Real vector spaces

Idea: work out what we've used about **R^n** in our analysis thus far. Other things which aren't obviously the same as **R^n** satisfy these properties too, so our theorems and our geometric intuition apply to those too.

Definition:
A _real vector space_ is a set **V** equipped with two operations:
***** vector addition: given **v** and **w** in **V**, can form the sum **v+w**
***** scalar multiplication: given **k** in **R** and **v** in **V**, can form the product kv
such that
1. **v+w** is in **V** for all **v**, w in **V**
2. **v+w = w+v** all **v**,w
3. **u+(v+w) = (u+v)+w** all **u**,**v**,w
4. there exists an element 0 of **V ("the zero vector")** such that **v+0 = v** for all **v**
5. for each **v** in **V**, there exists an element -v of **V** such that **v+(-v**) **= 0**
6. kv is in **V** all **k** in **R** and **v** in **V**
7. **k(v+w) = kv+kw** all **k** in **R** and **v**,w in **V**
8. (**k+l**)v **= kv+lv** all **k**,**l** in **R** and **v** in **V**
9. (kl)v **= k(lv)** all **k**,**l** in **R** and **v** in **V**
10. 1v **= v** all **v** in **V**

We call the elements of **V** "vectors"

Examples:
R^n with usual vector **+** and *****.

If **A** is an nxn matrix, then the set of **v** in **R^n** such that **Av=0**, with usual addition and multiplication, is a real vector space.
Only 1, 4 and 5 are not immediate, but if **Av=0** and **Aw=0** then **A(v+w)=0** and **A(-v)=0**, and **A0=0**.

The set **C** of complex numbers, with (**a+bi)+(c+di)=(a+c)+(b+d)i** and **r(a+bi)=ra+rbi**. (We "forget" about multiplying together complex numbers!)

Mat_n with element-wise addition and scalar multiplication.

The space of sequences **R^\omega**: elements are infinite sequences **(a_0,a_1,a_2,...)**; **+** and ***** are co-ordinatewise.

The space of functions **R -> R**, with (**f+g)(x) = f(x)+g(x)** and (**kf)(x) = k(f(x))**

R^* = R without 0, with **+** being ***** and ***** being **^**.

Lemma:
Let **V** be a vector space

(i) There is a _unique_ element 0 \in V such that v+0=v for all v
(ii) Given v\in V, there is a _unique_ element -v\in V such that
v + -v = 0
so our notation is justified.

Proof:
(i) By Axiom 4, there is at least one such 0. If 0' is another, then
0' = 0'+0 = 0+0' = 0
(ii) By Axiom 5, there is at least one such -v. If w is another, then
w = w + 0 = w + (v + -v) = -v + (v + w) = -v + 0 = -v

Lemma:
Let V be a vector space, let v\in V.
(i) 0v = 0
(ii) (-1)v = -v

Proof:
(i) w + 0v = w + 0v + 0v + -(0v) = w + (0+0)v + -(0v) = w + 0v + -(0v) = w
(ii) (-1)v + v = (-1)v + 1v = (-1+1)v = 0v

Subspaces:

Definition:
A subset U of a vector space V is a _subspace_ iff U is a vector space when equipped with the addition and scalar multiplication inherited from V.

Theorem:
A non-empty subset U of a vector space V is a subspace iff U is closed under addition and scalar multiplication i.e. iff for all u,v\in U and k\in\R, u+v \in U and ku \in U.

Proof:
=>: clear
<=: Axioms 2,3,7,8,9,10 hold because they hold of V.
We are assuming axioms 1 and 6.
Axioms 4 and 5 follow:
let u\in U, then 0u = 0 \in U, and (-1)u = -u\in U.

Examples:
Solutions to Ax=0 in R^n

Let F(R) be the space of functions R->R, and let C^n(R) be the subset of those functions f which are n-times continuously differentiable (f^(n) exists and is continuous). Then C^n(R) is a subspace of F(R). Note also that C^{n+1}(R) is a subspace of C^n(R).

Solutions to linear differential equations:
e.g. the subset of C^2(R) consisting of those f satisfying
f'' + f = 0
is a subspace.

Remark: If U and W are subspaces of a vector space V, then so is their intersection U \cap W.

Definition: The _span_ of a finite subset S = {u_1,...,u_n} of a vector space V is the set span(S) of linear combinations: the set of those elements of V which can be written as
k_1 u_1 + ... + k_n u_n
for some k_1,...,k_n \in R.

We say S _spans_ V iff span(S)=V.

Lemma: span(S) is a subspace of V.
Any subspace of V containing S contains span(S).

Proof: (k_1 u_1 + ... + k_n u_n) + (k_1' u_1 + ... + k_n' u_n) =
(k_1+k_1')u_1 + ... + (k_n+k_n')u_n

-(k_1 u_1 + ... + k_n u_n) = ((-k_1) u_1 + ... + (-k_n) u_n)

Example: in R^3, any line through the origin is spanned by any non-0 vector on the line;
any plane through the origin is spanned by any two non-colinear vectors on the plane.
Three (or more) vectors lying on no common plane span the whole of R^3.

Remark: in R^n, a set of n vectors {u_1,...,u_n} spans R^n
iff for every b\in R^n, the equation x_1u_1,...,x_nu_n=b

has a solution for $x_1,\dots,x_n \in R$
iff for every $b \in R^n$, the equation $Ax=b$ has a solution for $x \in R^n$
where $A = [u_1 \dots u_n]$
iff A is invertible.

In other words, A is singular iff $u_1,\dots u_n$ are in a proper subspace of R^n.

Linear independence

Let V be a vector space.

Definition:
A finite subset $\{v_1,\dots,v_n\}$ of V is _linearly independent_ if the only solution for $k_1,\dots,k_n$ to
$$k_1 v_1 + \dots + k_n v_n = 0$$
is the trivial solution $k_1 = \dots = k_n = 0$.

(Linguistic note: We often drop the set theoretic notation, and say "$v_1,\dots,v_n$ are linearly independent". This is "plural predication" - we're making a claim of how the vectors relate to each other, not merely claiming that each vector satisfies something.)

Theorem:
$v_1,\dots,v_n$ are linearly dependent iff one of them is in the span of the others.

Proof:
$\Rightarrow$: say $k_1 v_1 + \dots + k_n v_n = 0$ with at least one of the k_i not zero.
Reordering, we may assume $k_1 \neq 0$.
Then $v_1 = (k_2/k_1 v_2 + \dots + k_n/k_1 v_n) \in \text{span}(\{v_2,\dots,v_n\})$.

$\Leftarrow$: Reordering, say $v_1 \in \text{span}(\{v_2,\dots,v_n\})$.
So say $v_1 = k_2 v_2 + \dots + k_n v_n$.
Then $1\, v_1 + (-k_2) v_2 + \dots + (-k_n) v_n = 0$.
$1 \neq 0$, so $v_1,\dots,v_n$ are linearly dependent.

Remark:
$v_1 \in \text{span}(\{v_2,\dots,v_n\})$ iff $\text{span}(\{v_1,\dots,v_n\})=\text{span}(\{v_2,\dots,v_n\})$.
So linear dependence $\Leftrightarrow$ "redundancy" - we didn't need so many vectors to span the subspace.

Examples:
$\{v\}$ is linearly dependent iff $v=0$

$\{v,w\}$ is linearly dependent iff v and w are colinear.

In R^3: three vectors u,v,w are linearly dependent iff they lie on a common plane.

In $F(R)$:
$\sin(x)$ and $\cos(x)=\sin(x+\pi/2)$ are linearly independent
but e.g. $\sin(x),\cos(x),\sin(x+\pi/3)$ are linearly dependent,
since
$$\sin(x+\pi/3) = \cos(\pi/3)\sin(x) + \sin(\pi/3)\cos(x) = 0.5 \sin(x) + \sqrt{3/2} \cos(x)$$
similarly $\sin(x+a)\in\text{span}(\sin(x), \cos(x))$ for any a.

(note that $\sin(x+a)$ is a solution to $f''+f=0$...)

In R^n, n vectors $v_1,\dots,v_n$ are linearly dependent
iff $k_1 v_1 + \dots + k_n v_n = 0$ has a non-trivial solution
iff $Ax = 0$ has a non-trivial solution where $A = [v_1 \dots v_n]$
iff $\det(A) = 0$.

(recall signed volume interpretation of det)

Any $v_1,\dots,v_{n+1} \in R^n$ must be linearly dependent:
If $v_1,\dots,v_n$ are linearly dependent, then so are $v_1,\dots,v_{n+1}$.
Else, $A = [v_1 \dots v_n]$ is non-singular, so $\text{span}(\{v_1,\dots,v_n\}) = R^n$. But then $v_{n+1} \in \text{span}(\{v_1,\dots,v_n\})$.

Co-ordinates and bases

=====================

Definition: a subset S = {v_1,...,v_n} of a vector space V is a _basis_
if it is linearly independent and spans V (meaning span(S)=V).

e.g. in R^3: four or more vectors are necessarily linearly dependent. Two or fewer can't span R^3. So bases consist of three vectors. v_1,v_2,v_3 form a basis of R^3 iff they lie on no common plane through 0.

In R^n: If we have more than n vectors, we saw they must be linearly dependent. If we have exactly n, v_1,...,v_n, we saw that they span iff det([v_1 ... v_n]) is non-singular iff they are LI. If we have fewer than n, v_1,...,v_k, then they do not span R^n, since span({v_1,...,v_k}) = span({v_1,...,v_k,0,...,0}).

So bases of R^n are precisely of the form {v_1,...,v_n} with [v_1 ... v_n] non-singular.

Standard basis of R^n: e_i := (0,...,0,1,0,...,0)^T (so [e_1 ... e_n]=I).

Other examples:

Standard basis of M_22

Standard basis of P_n = {polynomials/R of deg <= n}

f''+f=0: sin, cos form a basis (Fact)

In R^n: let A be an nxn matrix. We saw that A is diagonalisable iff there is an invertible nxn matrix P whose columns are e-vectors of A.
So A is invertible iff R^n has a basis consisting of e-vectors of A.

<R_{>0};*,(.^r)_r>

Getting a basis out of a spanning set:
Suppose S={v_1,...,v_n} spans V. Then some subset of S is a basis:

If S is linearly independent, then S is a basis. Otherwise, some v_i is in the span of the rest. Let S' be the result of removing v_i from S. Then S' still spans V. If S' is linearly independent, S' is a basis and we're done. Otherwise, again we can throw something out to get a smaller spanning set. Keep going until we get a linearly independent spanning set.

Definition: a finite sequence (v_1,...,v_n) of vectors in a vector space V is an _ordered basis_ of V if the set {v_1,...,v_n} is a basis of V.

Theorem:
If B=(v_1,...,v_n) is an ordered basis of a vector space V, then any w \in V can be _uniquely_ expressed as a linear combination of v_1,...,v_n; i.e. there is a _unique_ choice of reals k_1,...,k_n such that
w = k_1 v_1 + ... + k_n v_n

Definition:
We call this unique (k_1,...,k_n)\in R^n the _co-ordinate vector_ of w with respect to B, and we write w_B = (k_1,...,k_n).

Proof:
Existence by spanning.

Uniqueness:
Suppose k_1 v_1 + ... + k_n v_n = w = k_1' v_1 + ... + k_n' v_n.
Then (k_1 - k_1') v_1 + ... + (k_n - k_n') v_n = 0, so by linear independence, k_i - k_i' = 0 for all i. So k_i=k_i' for all i.

So a choice of ordered basis B=(v_1,...,v_n) "co-ordinatises" V - it gives us a way to refer to the elements of V using n-vectors of reals. In other words, it yields a 1-1 correspondence between V and R^n
(k_1,...,k_n) |-> k_1v_1 + ... + k_nv_n
w |-> w_B = [the unique (k_1,...,k_n) such that w = k_1v_1+...+k_nv_n]

Remark:
Taking co-ordinates maps the addition and scalar multiplication of V to those of R^n, i.e.
(v+w)_B = v_B + w_B
(kv)_B = kv_B

Definition:
The _length_ of a basis is the number of vectors in it (so if $B=\{v_1,...,v_n\}$, then B has length n).

Theorem:
If B and B' are bases for a vector space V, then they have the same length.

Definition:
If V has a finite basis, the _dimension_ of V, $\dim(V)$, is the length of any basis.

If V has no finite basis (equivalently: has no finite spanning set), we say V has infinite dimension.

Examples:
$\dim(R^n) = n$
$\dim(P_n) = n+1$
$P_\infty :=$ vector space of all polynomials (of any degree) has infinite dimension.

Proof:
Say $B = (v_1,...,v_n)$ and $B' = (w_1,...,w_m)$ are ordered bases for V.
Then $(w_{1B},...,w_{mB})$ is an ordered basis for R^n.
But we saw that any basis of R^n must have length n, so $m=n$.

Upshot:
All real vector spaces of dimension n are essentially the same: they are all "disguised forms" of R^n. To penetrate the disguise, we need a basis (and there may not be a unique natural choice).

If $B=(v_1,...,v_n)$ and $B'=(v'_1,...,v'_n)$ are ordered bases for an n-dimensional vector space V, then given $w \in V$ we have two corresponding elements of R^n, w_B and $w_{B'}$. How do they relate to each other?

Theorem:
Let $P_{B\to B'}$ be the matrix whose ith column is $(v_i)_{B'}$.
Then for any w, $P_{B\to B'} w_B = w_{B'}$, and $P_{B\to B'}$ is the unique matrix with this property.

Definition:
This matrix $P_{B\to B'}$ is called the _change of basis_ matrix from B to B'.

Proof:
$P_{B\to B'} (v_i)_B = P_{B\to B'} e_i = (v_i)_{B'}$
and $P_{B\to B'}$ is clearly unique with this property.

Now any $w \in V$ can be expressed as $w = k_1v_1 + ... + k_nv_n$, and
$P_{B\to B'} (k_1v_1 + ... + k_nv_n)_B =$
$P_{B\to B'} ((k_1v_1)_B + ... + (k_nv_n)_B) =$
$k_1 P_{B\to B'} (v_1)_B + ... + k_n P_{B\to B'} (v_n)_B =$
$k_1 (v_1)_{B'} + ... + k_n (v_n)_{B'} =$
$((k_1v_1)_{B'} + ... + (k_nv_n)_{B'}) =$
$(k_1v_1 + ... + k_nv_n)_{B'}$.

Example:
plane

Remark:
Given ordered bases B,B',B'',
for any w, $P_{B'\to B''} P_{B\to B'} w_B = P_{B'\to B''} w_{B'} = w_{B''}$.
So $P_{B'\to B''} P_{B\to B'} = P_{B\to B''}$.

In particular, $P_{B'\to B} P_{B\to B'} = P_{B\to B} = I$.
So $P_{B\to B'}$ is invertible, and its inverse is $P_{B'\to B}$.

Example:
Let $E=(e_1,...,e_n)$ be the standard basis for R^n.

If $B=(v_1,...,v_n)$ is an ordered basis for R^n, then
$P_{B\to E} = [\ v_1\ ...\ v_n\ ]$.
(since $(v_i)_E = v_i$!)

Let A be an nxn matrix.
Let $A' = P_{E\to B} A P_{B\to E}$.
Then for any $w \in R^n$,
$A' w_B = P_{E\to B} A w_E = P_{E\to B} Aw = (Aw)_B$.

e.g. if $B=(v_1,...,v_n)$ is a basis of e-vectors for A, say $Av_i = l_iv_i$.
Let $P := P_{B\to E} = [v_1\ ...\ v_n]$

then D := P_{E->B} A P_{B->E} = P^-1 A P
is diagonal:
D e_i = D (v_i)_B = (A v_i)_B = (l_i v_i)_B = l_i (v_i)_B = l_i e_i.

Diagonalisability again:
For each \mu, the \mu-e-space of A is a subspace of R^n
since it is the solutions of (A-\mu I)x = 0.

geometric multiplicity of \mu = dim({\mu-e-space})

algebraic multiplicity of \mu = power of (\lambda-\mu) in \xi_A(\lambda)

Fact: geometric multiplicity <= algebraic multiplicity

Fact: if we put together bases for the e-spaces, the resulting set is
linearly independent.

So:
nxn A is diagonalisable in the reals
iff the sum of the dimensions of the eigenspaces is n
iff (i) \xi_A has n real zeroes
and (ii) for each e-value, the geometric multiplicity is equal
to the algebraic multiplicity.

(Note: if we work in C, (i) comes for free)

Example:
(3 1 0)
(0 3 0)
(0 0 0)
is not diagonalisable, since the 3-e-space is span({(1,0,0)}), so the
geometric multiplicity of 3 is 1, but \xi(l) = l(l-3)^2 so the
algebraic multiplicity of 3 is 2.

Orthogonal bases and Gram-Schmidt

A subset S={v_1,...,v_s} of R^n is _orthogonal_ if v_i.v_j=0 whenever i!=j.
S is _orthonormal_ if also ||v_i|| = 1, i.e. v_i.v_i=||v_i||^2=1.
An orthonormal set is always linearly independent:
if k_1v_1 + ... + k_sv_s = 0 then
k_i = (k_1v_1 + ... + k_sv_s).v_i = 0

If v_1,...,v_s are orthogonal and non-zero, then {v_1/||v_1||,...,v_s/||v_s||}
is orthonormal. So {v_1,...,v_s} is also LI.

Let W <= R^n be a subspace.
v\in R^n is _orthogonal to W_ if v.w=0 for all w\in W.
Suppose {w_1,...,w_k} is an orthonormal basis for W.
Let proj_W v := (v.w_1)w_1 + ... + (v.w_k)w_k \in W.
Note: proj_w v = proj_{span(w)} v
Then v - proj_W v is orthogonal to W.
(Proof: since {w_1,...,w_k} spans W, enough to see that
(v - proj_W v).w_i = 0. But indeed
(v - proj_W v).w_i = v.w_i - (v.w_1)w_1.w_i - ... - (v.w_k)w_k
= v.w_i - v.w_i
= 0)

Gram-Schmidt:
Let {u_1,...,u_k} be a basis for a subspace U of R^n.
We find an orthonormal basis {w_1,...,w_k} for U as follows:

Let w'_1 := u_1, let w_1 = w'_1 / ||w'_1||.

Let w'_2 := u_2 - proj_{W_2} (W_2 := span({w_1}))
= u_2 - (u_2.w_1)w_1
and w_2 := w'_2 / ||w'_2||.
(w'_2 != 0, since u_2 \notin W_2 = span({u_1}))

At the ith stage, we have an orthonormal set {w_1,...,w_{i-1}}, which is an
orthonormal basis for W_i := \span({w_1,...,w_{i-1}}).

```
Let w'_i := u_i - proj_{W_i}        (W_i := span({w_1,...,w_{i-1}))
          = u_i - (u_i.w_1)w_1 - ... - (u_i.w_{i-1})w_{i-1}
and let w_i := w'_i / ||w'_i||.
    (w'_i != 0, since u_i \notin W_i = span({u_1,...,u_{i-1}}))

Example:
    Find an orthonormal basis for
        span({(1,-1,0,0), (0,1,-1,0), (0,0,1,-1)}) <= R^4

// Say something about inner product spaces


Example:
    Find an orthogonal basis for
        span({(1,1,0,0),(0,0,1,1),(0,1,1,0)})

    u_3 - proj_u_1 u_3 - proj_u_2 u_3
      = u_3 - (u_1.u_3) u_1/||u_1||^2 - (u_2.u_3) u_2/||u_2||^2
      = u_3 - u_1/2 - u_2/2
      = (-1/2, 1/2, 1/2, -1/2)

Row space, column space, null spaces
====================================

Let A be an nxm matrix. Consider the rows as elements of R^m and the columns
as elements of R^n.

The _row space_ of A is the span of the rows, a subspace of R^m.
The _column space_ is the span of the columns, a subspace of R^n.
The _null space_ is the space of solutions to Ax=0, a subspace of R^m.

Remark:
    The null space consists of the elements of R^n which are orthogonal to the
    row space.

Lemma: suppose A' is the result of applying some elementary row operations to A.
    Then:
        (i) rowSpace(A') = rowSpace(A)
        (ii) null(A') = null(A)
        (iii) if c_1,...,c_m and c_1',...,c_m' are the columns of A and A',
            then they satisfy the same linear relations, i.e.
            for any k_1,...,k_m \in R,
                k_1c_1' + ... + k_mc_m' = 0
            iff
                k_1c_1 + ... + k_mc_m = 0
Example:
    (1 1 2)
    (1 3 2)

/* Remark:
    If R is rref:
        * the non-zero rows form a basis for the row space
        * the columns containing a leading one form a basis for the column space

                    (0 1 0 2 0 3)
                    (0 0 1 4 0 5)
                    (0 0 0 0 1 6)
*/

Corollary:
    To find bases for the row, column and null spaces of a matrix A:
        row reduce, yielding rref R; then
        (i) a basis for rowSpace(A) = rowSpace(R) is given by the non-zero rows of R
        (ii) a basis for colSpace(A) is given by the columns of A which are in
            the same positions as the columns of R which have leading 1s,
        (iii) each column not containing a leading 1 contributes a vector
            towards a basis for null(A)=null(R) - namely the unique vector in the
            null space which has 1 in that position and 0 in the positions of the
            other columns which don't contain leading 1s.

                ( e.g. null space of
                    (1 0 2 0 3)
                    (0 1 4 0 5)
```

(0 0 0 1 6)
has basis { (-2,-4,1,0,0), (-3,-5,0,-6,1) })

Hence:
dim(rowSpace(A)) = dim(colSpace(A)) = m - dim(null(A))

Example:
Find a basis for W := span({(1,1,1,1), (1,2,3,4), (4,3,2,1)})
and a basis for the space of vectors orthogonal to W

Example:
Find a basis for span({(1,1,1), (1,2,3), (3,2,1)}) which is a subset of
{(1,1,1), (1,2,3), (3,2,1)}.

Review
======

dim n vector space with a choice of basis <~~> R^n with standard basis

Theorem:
Let A be an nxn matrix. Then the following are equivalent:
(a) rowSpace(A) = R^n
(b) colSpace(A) = R^n
(c) null(A) = {0}
(d) 0 is not an e-value of A
(e) Ax=0 has no non-trivial solutions
(f) for every b, Ax=b has exactly one solution
(g) A is invertible - exists A^-1 such that AA^-1 = I = A^-1A
(h) A is row equivalent to I
(i) A is a product of elementary matrices
(j) det(A)!=0
When these are true, we say A is _non-singular_.
When they are false, we say A is _singular_.

Proof:
(a) <=> (b) <=> (c): last lecture

(c) <=> (d) <=> (e): definitions

(e) <=> (f): solution set of Ax=b is a shift of the solution space of Ax=0

(f) => (g):
Let A^-1 have ith column the unique x_i such that Ax_i = e_i.
Then AA^-1 e_i = Ax_i = e_i, so the ith column of AA^-1 is e_i,
i.e. AA^-1=I.

Then for any x, (A A^-1) (A x) = (A x), so AA^-1 = I.

(g) => (f):
A^-1b is the only solution.

(h) <=> (a): Gaussian elimination

(i) <=> (h): r(A) = r(I)A

(j) <=> (h): e.r.o.s multiply determinants by non-zero scalars

Determinants: u.(vxw); signed volume

Solving Ax=lx

Diagonalising; powers

Complex numbers to find all the e-values

Example: V = span{sin(2x), cos(2x)}
B = (sin(2x), cos(2x)) ordered basis
dsin(2x)/dx = 2cos(2x); dcos(2x)/dx = -2sin(2x)
so
(df/dx)_B = A (f_B)
where A = (0 -2)
(2 0)
l^2+4=0

l=+−2i

D = diag(2i,−2i)

(1 −i)
(0 0)

2i−e−space **= span({cos(2x) + i sin(2x)}) = span({e^2ix})**
−2i−e−space **= span({cos(2x) − i sin(2x)}) = span({e^−2ix})**

As a _complex_ vector space (still taking **x** to be real)
B' = {e^2ix, **e^−2ix}** is another basis
and (**df/dx)_B' = D f_B**