Math UA 251

Section 4

Spring 2024

Christiana Mavroyiakoumou (cm4291@nyu.edu)

lecture 1

(read Topics in Mathematical Modeling by Tung, Chapter 1)

Fibonacci Numbers

Puzzle. A man puts a pair of rabbits in a room. How many pairs of rabbits can be produced from that pair in a year if we suppose that each month each pair reproduces a new pair which from the 2nd month on becomes productive?

Q. Find the number of pairs of rabbits n months after the 1st pair was introduced.

A We denote this quantity by F_n.

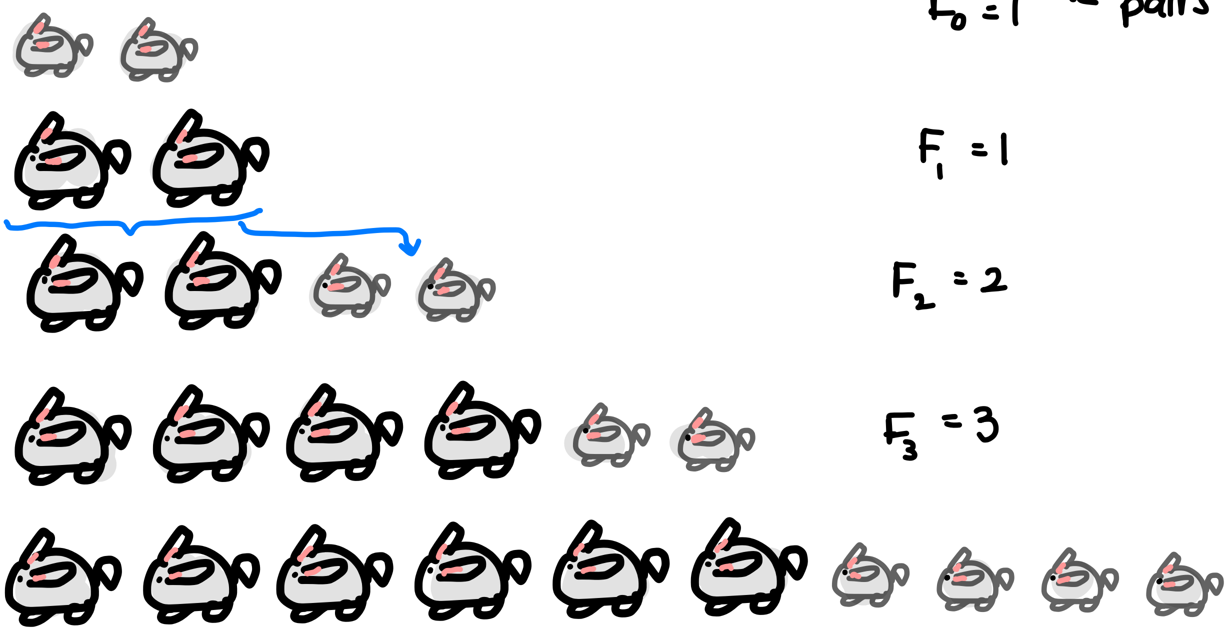
Month 0: $F_0 = 1$ ← # of pairs

Month 1: $F_1 = 1$

Month 2: $F_2 = 2$

Month 3: $F_3 = 3$

Month 4: $F_4 = 5$



Pattern: Any number in the sequence is always a sum of the two numbers preceding it. i.e.

$$\boxed{F_{n+2} = F_{n+1} + F_n} \quad \text{for } n = 0, 1, 2, 3, \ldots$$

But we can also use a recurrence relationship w/o detecting a pattern.

Let $F_n(k)$ be the number of k-month-old rabbit pairs at time n.

These will become (k+1)-month-old rabbits at time n+1.

$$F_{n+1}(k+1) = F_n(k)$$

The total number of pairs at time n+2 is equal to the number at n+1 plus the newborn pairs at n+2

$$(*) \qquad F_{n+2} = F_{n+1} + \underline{\text{new births at time } n+2}$$

= number of pairs that are at least one month old at n+1

$$= F_{n+1}(1) + F_{n+1}(2) + F_{n+1}(3) + F_{n+1}(4) + \dots$$

$$= F_n(0) + F_n(1) + F_n(2) + F_n(3) + \dots$$

$$= F_n$$

↑ one less month old the month before...

Thus (*) becomes

$$\boxed{F_{n+2} = F_{n+1} + F_n}$$

Mathematically it's a 2nd-order difference equation

To solve this we use as an Ansatz: $\boxed{F_n = \lambda^n}$

$$\lambda^{n+2} = \lambda^{n+1} + \lambda^n$$

$$\lambda^2 \cancel{\lambda^n} = \cancel{\lambda^n}(\lambda + 1)$$

$$\Rightarrow \lambda^2 = \lambda + 1$$

$$\Rightarrow \lambda^2 - \lambda - 1 = 0$$

$$\left(\lambda - \frac{1}{2}\right)^2 - \frac{1}{4} - 1 = 0$$ by completing the square

$$\left(\lambda - \frac{1}{2}\right)^2 - \frac{5}{4} = 0$$

$$\lambda = \frac{1}{2} \pm \frac{\sqrt{5}}{2}$$

So the two solutions are $\lambda_1 = \frac{1+\sqrt{5}}{2}$, $\lambda_2 = \frac{1-\sqrt{5}}{2} = -\frac{1}{\lambda_1}$

Thus λ_1^n, λ_2^n are both solutions. By the principle of linear superposition. the general solution is

$$F_n = \underset{\uparrow}{a}\lambda_1^n + \underset{\nearrow}{b}\lambda_2^n.$$

arbitrary constants but they can be determined from initial conditions.

e.g. if $F_0 = 1, F_1 = 1$

$$F_0 = 1 \Rightarrow a\underset{1}{\cancel{\lambda_1^0}} + b\underset{1}{\cancel{\lambda_2^0}} = 1 \Rightarrow \boxed{a+b=1} \Rightarrow b = 1-a$$

$$F_1 = 1 \Rightarrow a\lambda_1 + b\lambda_2 = 1$$

$$a\lambda_1 + (1-a)\lambda_2 = 1$$

$$a[\lambda_1 - \lambda_2] + \lambda_2 = 1$$

$$a\left[\frac{\cancel{1}+\sqrt{5}}{2} - \left(\frac{\cancel{1}-\sqrt{5}}{2}\right)\right] + \frac{1-\sqrt{5}}{2} = 1$$

$$a\left[\sqrt{5}\right] = 1 - \frac{1}{2} + \frac{\sqrt{5}}{2} = \frac{1}{2} + \frac{\sqrt{5}}{2}$$

$$a = \frac{1}{\sqrt{5}}\left(\frac{1}{2} + \frac{\sqrt{5}}{2}\right) \quad \text{and } b = 1 - a$$

$$= 1 - \frac{1}{\sqrt{5}}\left(\frac{1}{2} + \frac{\sqrt{5}}{2}\right)$$

$$= -\frac{1}{\sqrt{5}}\left(\frac{1}{2} - \frac{\sqrt{5}}{2}\right)$$

Thus plugging these into $F_n = a\lambda_1^n + b\lambda_2^n$ we obtain

$$(\dagger) \quad \boxed{F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}$$

note that the exponent has increased by 1.

<u>Exercise</u> Verify that even with the irrational number $\sqrt{5}$ in the expression, Eq $(\dagger)$ always yields whole number $1, 1, 2, 3, 5, 8, \ldots$ when n goes from $0, 1, 2, 3, 4, \ldots$

THE GOLDEN RATIO

The number $\lambda_1 = \frac{1+\sqrt{5}}{2}$ is known as the golden ratio. We denote it by Φ

It reflects the ideal proportions of nature.

It has some special properties:

$$\Phi \simeq 1.6180339887\ldots$$

$$\Phi^2 = 2.6180339887\ldots = \Phi + 1$$

$$\frac{1}{\Phi} = 0.6180339887\ldots = \Phi - 1$$

But these are not mysterious if we remember that Φ solves

$\Phi^2 = \Phi + 1$ (recall we found λ from solving $\lambda^2 = \lambda + 1$)

In terms of the golden ratio we can write the general solution as

$$F_n = a\Phi^n + b\left(-\frac{1}{\Phi}\right)^n$$

Since $\Phi > 1$, as $n \to \infty$ we have $F_n \to a\Phi^n$.

Thus the ratio of successive terms in the Fibonacci sequence approaches the Golden ratio:

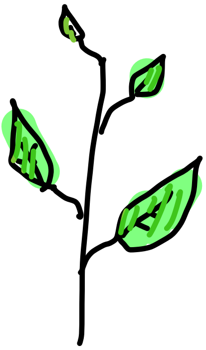
$$\frac{F_{n+1}}{F_n} \to \frac{a\Phi^{n+1}}{a\Phi^n} = \Phi = 1.6180339887\ldots \quad \text{as } n \to \infty$$

Phyllotaxis

Phyllotaxis is the study of leaf arrangements in plants.

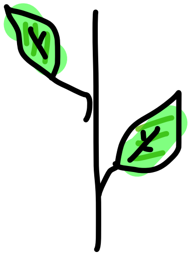
Fibonacci numbers are prevalent in the phyllotaxis of various trees, e.g in seed heads, pinecones, and sunflowers.

As the stem of a plant grows upward, leaves sprout to its side, with new leaves above the old ones

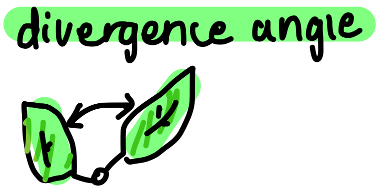


Q How are the new and old leaves arranged?
Is there a pattern?

The Bravais brothers (in 1837) discovered that a new leaf advances by the same angle from the previous leaf and that angle is ~ 137.5°.



top view



One could think that the divergence angle should be something simple like 180°. That would mean that the new leaf would be directly opposite from the older leaf, perhaps to provide balance for the plant.

However, if the plant has many leaves, then if this were the case for leaf 0 and leaf 1 then leaf 2 would be directly above leaf 0, blocking sun exposure and water absorption from rainfall.

ALSO BAD:

Any divergence angle which is an integer fraction of the circle, i.e $\frac{360^\circ}{m}$, $m \in \mathbb{Z}$ is not optimal for the plant

⇒ periodic arrangement

⇒ eventually some new leaves directly above some old leaves

GOOD

Replace the integer m by an irrational number — the more irrational the better. It turns out the Golden ratio $\Phi = 1.618\ldots$ is the best.

Divergence angle $= \frac{360°}{\Phi} = 222.5°$ which is the same as

$360 - 222.5 = \underbrace{137.5°}_{\downarrow}$ measuring from the other side.

Golden Angle

Definition: Phyllotactic ratio is the fraction of a circle through which a new leaf turns from the previous, older leaf.

So in this case the phyllotactic ratio is $\frac{1}{\Phi} = 0.618\ldots$

Since $\frac{1}{\Phi} > 0.5$, i.e. more than half of the circle we can measure the angle from the other direction $1 - \frac{1}{\Phi} \approx 0.382$

Recall that $\Phi^2 = \Phi + 1$

$$\frac{1}{\Phi} = \Phi - 1$$

$$\Rightarrow \quad 1 - \frac{1}{\Phi} = \frac{\Phi - 1}{\Phi} = \frac{\left(\frac{1}{\Phi}\right)}{\Phi} = \frac{1}{\Phi^2}$$

and we have already seen that as $n \to \infty$, $F_n \approx a\Phi^n$ Thus

$$\frac{F_n}{F_{n+2}} \approx \frac{a\Phi^n}{a\Phi^{n+2}} = \frac{1}{\phi^2}$$

where F_n is one of the Fibonacci numbers. The phyllotactic ratio is ratio of <u>every other</u> Fibonacci number. If one measures the angle in the other direction; ↻ instead of ↺ then one will detect a different set of Fibonacci numbers:

$$\text{phyllotactic ratio} = \frac{1}{\Phi} \approx \frac{F_n}{F_{n+1}}$$

The above arguments apply to plants with many leaves (actually, an infinite number of leaves) & with the assumption that the only determining factor for the arrangement of leaves in a plant is sun exposure

Lecture 2 (read Tung's book, Chapter 9)

Consider 2 or more interacting species. → Coupled set of nonlinear ODEs

NONLINEAR SYSTEM AND ITS LINEAR STABILITY

$$\left.\begin{aligned} \frac{dx}{dt} &= f(x,y) \\ \frac{dy}{dt} &= g(x,y) \end{aligned}\right]$$

$x(t), y(t)$ are the two interacting species
f, g are nonlinear functions of x and y

To retrieve info about the behavior of the system we do the following:

1. Find the equilibrium solutions x^* and y^* by solving the simultaneous eqns.:

$$\text{and} \quad \boxed{\begin{aligned} f(x^*, y^*) &= 0 \\ g(x^*, y^*) &= 0 \end{aligned}}$$

2. Determine if the equilibrium is <u>stable</u> or <u>unstable</u>

=> Small perturbations from eqm.

(a) Linearize the nonlinear equations about (x^*, y^*) ← the eqm sol^n

$$\begin{aligned} x(t) &= x^* + u(t) \\ y(t) &= y^* + v(t) \end{aligned}$$

This implies that $\frac{dx}{dt} = \frac{d}{dt}(x^* + u(t))$

$$= \underset{\substack{0 \\ \text{by def}^n}}{\cancel{\frac{dx^*}{dt}}} + \frac{du}{dt} \quad \Rightarrow \quad \frac{dx}{dt} = \frac{du}{dt}$$

Similarly, $\frac{dy}{dt} = \frac{dv}{dt}$.

(b) Expand f and g about the eqm in a Taylor series

$$f(x,y) = \overset{0 \text{ by defn}}{\cancel{f(x^*,y^*)}} + \frac{\partial f}{\partial x}(x^*,y^*)\underbrace{(x-x^*)}_{=u} + \frac{\partial f}{\partial y}(x^*,y^*)\underbrace{(y-y^*)}_{=v}$$

$$+ \text{h.o.t}$$

$$\cong a_{11}u + a_{12}v$$

Where $\frac{\partial f}{\partial x} =: a_{11}$, $\frac{\partial f}{\partial y} =: a_{12}$

NOTE The process of dropping the higher order terms (i.e the nonlinear ones) is called LINEARIZATION.

Valid only if we want to study the behavior of the solution close to (x^*, y^*)

Similarly $g(x,y) \cong a_{21}u + a_2 v$, with $\frac{\partial g}{\partial x}(x^*,y^*) = a_{21}$, $\frac{\partial g}{\partial y} = a_{22}$

3. Coupled linear system:

$$\left.\begin{aligned} \frac{du}{dt} &= a_{11}u + a_{12}v \\ \frac{dv}{dt} &= a_{21}u + a_{22}v \end{aligned}\right] \rightarrow \frac{d}{dt}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

$$\Rightarrow \frac{d\vec{u}}{dt} = A\vec{u}$$

4 For linear equations with constant coefficients we have as

Ansatz: $u(t) = u_0 e^{\lambda t}$

$v(t) = v_0 e^{\lambda t}$

Subst. into $\frac{d\vec{u}}{dt} = A\vec{u}$ to get:

$$\begin{pmatrix} \lambda u_0 \cancel{e^{\lambda t}} \\ \lambda v_0 \cancel{e^{\lambda t}} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u_0 \cancel{e^{\lambda t}} \\ v_0 \cancel{e^{\lambda t}} \end{pmatrix}$$

$$\begin{pmatrix} \lambda u_0 \\ \lambda v_0 \end{pmatrix} = \begin{pmatrix} a_{11} u_0 + a_{12} v_0 \\ a_{21} u_0 + a_{22} v_0 \end{pmatrix}$$

Or, equivalently,

$$\begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

To have nontrivial solutions we must have

$$\det \begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} = 0$$

$$\lambda^2 - \underbrace{(a_{11} + a_{22})}_{p} \lambda + \underbrace{a_{11} a_{22} - a_{12} a_{21}}_{q} = 0$$

So we can rewrite this as $\boxed{\lambda^2 - p\lambda + q = 0}$

where $p \equiv Tr(A)$ and $q = \det(A)$

trace & determinant of matrix A, respectively

Solving the quadratic equ we get that the eigenvalues are

$$\lambda_1 = \frac{p}{2} + \frac{\sqrt{p^2 - 4q}}{2}, \quad \lambda_2 = \frac{p}{2} - \frac{\sqrt{p^2 - 4q}}{2}$$

p, q determine the STABILITY of the system.

- If $q<0 \Rightarrow \lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1>0$, $\lambda_2<0$

  Eqm is a saddle point $\Rightarrow$ unstable.

  [General solution is $\underset{\text{grows}}{c_1 e^{\lambda_1 t}} + \underset{\text{decays}}{c_2 e^{\lambda_2 t}}$]

- If $0<q<p^2/4 \Rightarrow \lambda_1, \lambda_2 \in \mathbb{R}$ with the same sign.

  For $p<0 \Rightarrow \lambda_1, \lambda_2<0$ STABLE NODE

  For $p>0 \Rightarrow \lambda_1, \lambda_2>0$ UNSTABLE NODE

- If $q>\frac{p^2}{4} \Rightarrow \lambda_1, \lambda_2 \in \mathbb{C} \Rightarrow$ oscillations.

  [$\lambda = a+ib$

  $e^{\lambda t} = e^{at} e^{ibt}$

  $= e^{at}(\underbrace{\cos bt + i \sin bt}_{\text{from Euler's identity}})$]

  Whether the amplitude of the oscillation will increase or decrease in t depents on the sign of p.

  For $p<0 \Rightarrow$ STABLE SPIRAL

  For $p>0 \Rightarrow$ UNSTABLE SPIRAL

  For $p=0 \Rightarrow$ CENTER

The general solution is $\vec{u} = \vec{u}_0^{(1)} e^{\lambda_1 t} + \vec{u}_0^{(2)} e^{\lambda_2 t}$ where $\vec{u}_0^{(1)}, \vec{u}_0^{(2)}$ are constant vectors. $\vec{u}_0^{(1)}$ is known as the eigenvector corresponding to the eigenvalue λ_1.

If $\lambda = a+ib$ we saw above that we obtain

$$e^{\lambda t} = e^{at}(\cos(bt) + i \sin(bt))$$

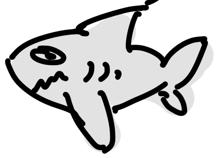
So if either λ_1 or λ_2 have a positive real part then the general solution will grow in time (so the origin $u=0, v=0$ is unstable).

However, the origin is stable only if <u>both</u> λ_1, λ_2 have a negative real part.

LOTKA-VOLTERRA PREDATOR-PREY MODEL

$$\boxed{\begin{aligned}\frac{dx}{dt} &= rx - axy\\ \frac{dy}{dt} &= bxy - ky\end{aligned}}$$

$x(t)$ = <u>prey</u> population density (e.g small 🐟)

$y(t)$ = <u>predator</u> population density (e.g. )

- small fish 🐟 eat <u>algae</u> and grow at a per capita rate $\left(\frac{dx}{dt}/r\right)$ of r
- Small fish are eaten by the sharks  and so their population density decreases at a per capita rate which is proportional to y

$$\underbrace{\frac{1}{x}\frac{dx}{dt}}_{\substack{\text{per capita}\\ \text{rate}}} = \underset{\substack{\text{algae}\\ \text{helps small fish}\\ \text{grow in population}}}{r} - \overset{\text{constant of proportionality}}{a}\,y \leftarrow \text{population of sharks}$$

- The predators (sharks) will die off without food. If $x=0$, $\frac{1}{y}\frac{dy}{dt}$ decreases at rate k
- In the presence of prey (small fish), the population of predators grows at a per capita rate of bx. This is proportional to the amount of food available.

$$\underbrace{\frac{1}{y}\frac{dy}{dt}}_{\text{per capita rate}} = \underbrace{bx}_{\text{feeding on prey}} - k$$

dying off w/o food

constant of proportionality

Lecture 3

LINEAR ANALYSIS

Consider the equilibria (x^*, y^*).

Set $\frac{dx}{dt} = 0$ & $\frac{dy}{dt} = 0$.

$$rx^* - ax^*y^* = 0 \Rightarrow x^*(r - ay^*) = 0 \Rightarrow \boxed{x^* = 0,\ y^* = \frac{r}{a}}$$

$$bx^*y^* - ky^* = 0$$

Subst. $x^* = 0$ in the 2nd eqn we obtain $y^* = 0$. Thus, one of the eqm pts is $(x_1^*, y_1^*) = (0, 0)$

The 2nd one comes from subst. $y^* = \frac{r}{a}$ into $y^*(bx^* - k) = 0$ to get $x^* = \frac{k}{b}$. Thus $(x_2^*, y_2^*) = \left(\frac{k}{b}, \frac{r}{a}\right)$

STABILITY OF THE EQUILIBRIA

★ Perturb slightly by the amount (u, v). I.e.

$$x(t) = x^* + u(t)$$
$$y(t) = y^* + v(t)$$

and then follow the method previously desribed w/ Taylor series expansions and the computation of eigenvalues/eigenvectors.

Alternatively, for $(x^*, y^*) = (x_1^*, y_1^*) = (0,0)$, we see that by substituting $\begin{bmatrix} x(t) = \cancelto{0}{x^*} + u(t) \\ y(t) = \cancelto{0}{y^*} + v(t) \end{bmatrix}$ into the system of ODEs we get

$$LHS_1 = \frac{dx}{dt} = \frac{du}{dt} \quad \text{and} \quad RHS_1 = rx - axy = ru - auv$$

$$LHS_2 = \frac{dy}{dt} = \frac{dv}{dt} \quad \text{and} \quad RHS_2 = bxy - ky = buv - kv$$

Therefore the governing eqns for the evolution of the perturbations is

$$\frac{du}{dt} = ru - auv$$

$$\frac{dv}{dt} = buv - kv$$

If the perturbations are small, we drop the quadratic terms to get

$$\frac{du}{dt} \simeq ru$$

$$\frac{dv}{dt} \simeq -kv$$

This is a linear system of ODEs: $\frac{d}{dt}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & -k \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$

Computing the eigenvalues we get

$$\det\begin{pmatrix} r-\lambda & 0 \\ 0 & -k-\lambda \end{pmatrix} = 0 \Rightarrow \lambda = r, \lambda = -k$$

$\boxed{\text{unstable saddle}}$

$$u(t) = u(0)e^{rt}$$
$$v(t) = v(0)e^{-kt}$$

Interpretation :

- A small increase from (0,0) will lead to an exponential growth in the prey (predators very few, algae abundant)
- A small increase in predators will not lead to an increase in the predator population. Actually they will die of starvation because the prey are very few.

→ The eqm (0,0) is still UNSTABLE because one of the populations does not stay low when perturbed.

Near the 2nd equilibrium $(x_2^*, y_2^*) = \left(\frac{k}{b}, \frac{r}{a}\right)$ we have

$$x(t) = x_2^* + u(t) = \frac{k}{b} + u(t)$$
$$y(t) = y_2^* + v(t) = \frac{r}{a} + v(t)$$

$$LHS_1 = \frac{dx}{dt} = \frac{du}{dt}, \quad RHS_1 = rx - axy = r\left(\frac{k}{b} + u\right) - a\left(\frac{k}{b} + u\right)\left(\frac{r}{a} + v\right)$$

$$= \cancel{r\frac{k}{b}} + \cancel{ru} - \cancel{a\frac{k}{b}\frac{r}{a}} - \cancel{au\frac{r}{a}} - a\frac{k}{b}v - auv$$

$$= -a\frac{k}{b}v - auv$$

Thus if we retain only the linear terms, we have

$$\boxed{\frac{du}{dt} \approx -a\left(\frac{k}{b}\right)v.}$$

Similarly, we have

$$LHS_2 = \frac{dy}{dt} = \frac{dv}{dt} . \quad RHS_2 = bxy - ky$$

$$= b\left(\frac{k}{b} + u\right)\left(\frac{r}{a} + v\right) - k\left(\frac{r}{a} + v\right)$$

$$= \cancel{k\frac{r}{a}} + bu\frac{r}{a} + \cancel{kv} + buv - \cancel{k\frac{r}{a}} - \cancel{kv}$$

$$= b\left(\frac{r}{a}\right)u + buv$$

Thus, retaining only the linear terms again. we have

$$\boxed{\frac{dv}{dt} \approx b\left(\frac{r}{a}\right)u}$$

We again have a linear system of equations

$$\frac{d}{dt}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 & -a\left(\frac{k}{b}\right) \\ b\left(\frac{r}{a}\right) & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

Computing the eigenvalues/eigenvectors we have

$$(-\lambda)(-\lambda) + \cancel{a}\left(\frac{k}{\cancel{b}}\right)\cancel{b}\left(\frac{r}{\cancel{a}}\right) = 0$$

$$\lambda^2 + kr = 0$$

$$\lambda = \pm i\sqrt{kr} \quad \Rightarrow \quad \underline{\text{CENTER}}$$

$$u(t) = c_1 \cos(\sqrt{kr}\, t) + c_2 \sin(\sqrt{kr}\, t)$$

and since $v(t), u(t)$ are related through $\boxed{\frac{du}{dt} = -a\left(\frac{k}{b}\right)v}$

we have

$$\frac{du}{dt} = \frac{d}{dt}\left[c_1 \cos(\sqrt{kr}\, t) + c_2 \sin(\sqrt{kr}\, t)\right]$$

$$= -c_1\sqrt{kr}\,\sin(\sqrt{kr}\, t) + c_2\sqrt{kr}\,\cos(\sqrt{kr}\, t)$$

$$= -a\frac{k}{b}v$$

$$\Rightarrow \quad v(t) = -\frac{b}{ak}\sqrt{kr}\left[-c_1\sin(\sqrt{kr}\, t) + c_2\cos(\sqrt{kr}\, t)\right]$$

$$= \frac{b}{a}\sqrt{\frac{r}{k}}\left[c_1\sin(\sqrt{kr}\, t) - c_2\cos(\sqrt{kr}\, t)\right]$$

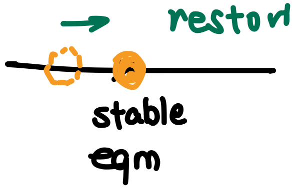
The solution $(u(t), v(t))$ is oscillatory with period $\frac{2\pi}{\sqrt{kr}}$

(Chapter 3 of Classical dynamics of particles & systems)

Oscillations – Simple harmonic oscillator

Consider the oscillatory motion of a particle constrained to move in one dimension.

Assume that a position of stable equilibrium exists for the



→ restoring force acts to take it back to its original position if it is displaced

Here we will consider only cases in which the restoring force F is a function only of the displacement : $F = F(x)$.

We assume that $F(x)$ possesses continuous derivatives of all orders so that the function can be expanded in a Taylor series :

$$F(x) = \underset{\substack{\uparrow \\ \text{value of } F(x) \text{ at the origin } (x=0)}}{F_0} + x\left(\frac{dF}{dx}\right)_0 + \frac{x^2}{2!}\left(\frac{d^2F}{dx^2}\right)_0 + \frac{x^3}{3!}\left(\frac{d^3F}{dx^3}\right)_0 + \dots$$

and $\left(\frac{d^nF}{dx^n}\right)_0$ = value of the n^{th} derivative at the origin.

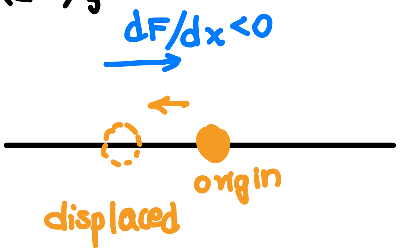
Since the origin is defined to be the equilibrium point, the restoring force F_0 must vanish. $\Rightarrow$ $\boxed{F_0 = 0}$

We focus on cases where the particle's displacements are small and so we neglect terms involving x^2 or higher powers of x.

Thus $\boxed{F(x) = -kx}$ (approximate relation), where we have subst. $k \equiv -\left(\frac{dF}{dx}\right)_0$.

HOOKE'S LAW

The restoring force is always directed toward the eqm position (i.e. the origin) and so the derivative $\left(\frac{dF}{dx}\right)_0 < 0$ and $\Rightarrow$ $\boxed{k > 0}$



Elastic deformations : As long as the displacements are small & the elastic limits are not exceeded, a linear restoring force can be used

stretched springs, elastic springs, bending beams, ...
obey HOOKE's law

In nature, almost always ⤳ damped oscillations resulting from friction

This damping can be counteracted if some mechanism supplies energy from an external source at a rate equal to that absorbed by the damping medium.
→ driven/forced oscillations.

SIMPLE HARMONIC OSCILLATOR

Newton's 2nd law of motion: $F = ma = m\ddot{x}$

Hooke's law: $F = -kx$

$$\left.\begin{array}{l} F = ma = m\ddot{x} \\ F = -kx \end{array}\right\} \Rightarrow -kx = m\ddot{x}$$

double dot on top of x denotes 2nd derivative wrt time,

$$\boxed{\ddot{x} = \frac{d^2x}{dt^2}}$$

If we define $\boxed{\omega_0^2 = \frac{k}{m}}$ then we have

$$-kx = m\ddot{x}$$

$$\ddot{x} + \frac{k}{m}x = 0$$

$$\boxed{\ddot{x} + \omega_0^2 x = 0}$$

This is a 2nd order ordinary differential equation (ODE) with constant coeff.
Its solution can be found using the characteristic equation $r^2 + \omega_0^2 = 0$

$$r = \pm i\omega_0$$

which means it can be expressed as either

$$\boxed{x(t) = A\sin(\omega_0 t - \delta)}$$

OR

$$x(t) = A\cos(\omega_0 t - \phi)$$

sinusoidal behavior of the displacement of the simple harmonic oscillator

where the phases δ, ϕ differ by $\frac{\pi}{2}$.

Relationship between total energy of the oscillator and the amplitude of the motion.

Kinetic energy

$$T = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{x}^2 = \frac{1}{2}m\left(A\omega_0\cos(\omega_0 t - \delta)\right)^2$$

$$= \frac{1}{2}mA^2\omega_0^2\cos^2(\omega_0 t - \delta)$$

but $\omega_0^2 = \frac{k}{m}$ and so $T = \frac{k}{2}A^2\cos^2(\omega_0 t - \delta)$

The potential energy can be obtained by calculating the work required to displace the particle a distance x.

Amount of work dW needed to move the particle a distance dx against the restoring force F is

$$dW = -F\,dx \quad (\text{force} \times \text{distance} = \text{work})$$
$$= kx\,dx \quad (F = -kx)$$

Integrating from 0 to x and setting the work done on the particle equal to the potential energy, gives

$$U = \frac{1}{2}kx^2$$

Thus $U = \frac{1}{2}k\left(A\sin(\omega_0 t - \delta)\right)^2 = \frac{1}{2}kA^2\sin^2(\omega_0 t - \delta)$

Therefore, if we combine the kinetic & potential energies to get the total energy E, we obtain

$$E = T + U = \frac{k}{2}A^2\cos^2(\omega_0 t - \delta) + \frac{1}{2}kA^2\sin^2(\omega_0 t - \delta)$$
$$= \frac{1}{2}kA^2\Big(\underbrace{\cos^2(\omega_0 t - \delta) + \sin^2(\omega_0 t - \delta)}_{1}\Big)$$
$$= \frac{1}{2}kA^2$$

Thus $\boxed{E = \frac{1}{2}kA^2}$ implies that the total energy is proportion to the square of the amplitude. E is independent of time ⤳ energy is conserved.

The period τ_0 of the motion is defined as the time interval between successive repetitions of the particle's position and direction of motion.

Recall $x(t) = A\sin(\omega_0 t - \delta)$ and since sine has a period of 2π

$$\omega_0 \tau_0 = 2\pi$$
$$\tau_0 = \frac{2\pi}{\omega_0} = \frac{2\pi}{\sqrt{\frac{k}{m}}}$$

← thus ω_0 represents the angular frequency of the motion

$$\Rightarrow \boxed{\tau_0 = 2\pi\sqrt{\frac{m}{k}}}$$

$$\omega_0 = 2\pi f_0 = \sqrt{\frac{k}{m}} \Rightarrow \underset{\substack{\uparrow \\ \text{frequency}}}{f_0} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{\tau_0}$$

Damped oscillations

Dissipative or frictional forces will eventually damp the motion to the point where the oscillations will cease.

⇒ We incorporate into the differential equation a term to represent the damping force.
It could be a function of the velocity or a higher time derivative of the displacement, e.g. $F_d = -bv \Rightarrow \boxed{F_d = -b\dot{x}}$

The parameter b must be positive for the force to be resisting

- $-b\dot{x}$ with $b<0$ would act to increase the speed instead of decreasing it as any resisting force must.

The ODE is now $\left.\begin{array}{l} F = m\ddot{x} \\ F = -kx - b\dot{x} \end{array}\right\} \Rightarrow \boxed{m\ddot{x} + b\dot{x} + kx = 0}$

Which we can rewrite as $\ddot{x} + \frac{b}{m}\dot{x} + \frac{k}{m}x = 0$

$$\boxed{\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0}$$

where we have defined $\boxed{\beta = \frac{b}{2m}}$ as the damping parameter and $\omega_0 = \sqrt{\frac{k}{m}}$ is as before the characteristic angular frequency in the absence of damping.

For this 2nd order ODE the characteristic equation is

$$r^2 + 2\beta r + \omega_0^2 = 0$$

and so if we solve for using the quadratic formula, we obtain

$$r = \frac{-2\beta \pm \sqrt{(2\beta)^2 - 4\omega_0^2}}{2} = -\frac{2\beta \pm \sqrt{4\beta^2 - 4\omega_0^2}}{2}$$

$$= -\beta \pm \sqrt{\beta^2 - \omega_0^2}$$

$$r_1 = -\beta + \sqrt{\beta^2 - \omega_0^2}$$

$$r_2 = -\beta - \sqrt{\beta^2 - \omega_0^2}$$

The general solution is

$$x(t) = Ae^{r_1 t} + Be^{r_2 t}$$

$$= e^{-\beta t}\left[Ae^{\sqrt{\beta^2 - \omega_0^2}\, t} + Be^{-\sqrt{\beta^2 - \omega_0^2}\, t}\right]$$

The 3 general cases of interest are

underdamping: $\omega_0^2 > \beta^2$

Critical damping: $\omega_0^2 = \beta^2$

overdamping: $\omega_0^2 < \beta^2$

Underdamped motion

We define $\omega_1^2 \equiv \omega_0^2 - \beta^2$ where $\omega_1^2 > 0$

Since the general solution is $x(t) = e^{-\beta t}\left[Ae^{\sqrt{\beta^2 - \omega_0^2}\, t} + Be^{-\sqrt{\beta^2 - \omega_0^2}\, t}\right]$ the exponent in the exponential function is imaginary and the solution becomes

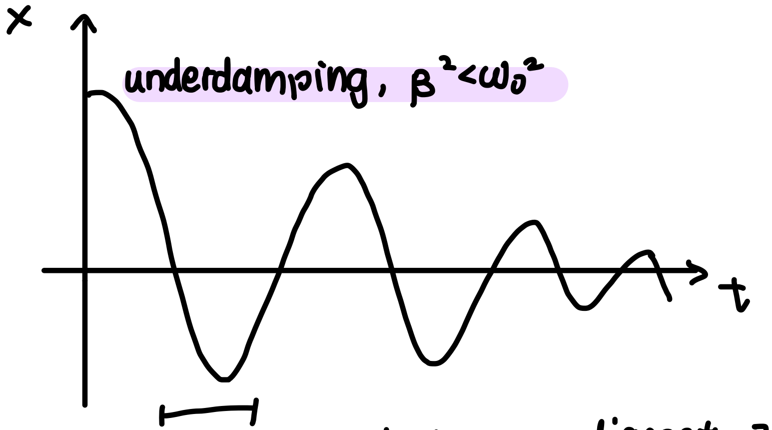
$$x(t) = e^{-\beta t}\left[Ae^{i\omega_1 t} + Be^{-i\omega_1 t}\right]$$

We can rewrite this as

$$x = Ce^{-\beta t}\cos(\omega_1 t - \delta)$$

SHOW THIS AS AN EXERCISE

ω_1 = angular frequency of the damped oscillator



T_1 = time between adjacent zero x-axis crossings

$$\omega_1 = \frac{2\pi}{(2T_1)} \leftarrow \text{"period"} \quad \Rightarrow \quad \boxed{\omega_1 = \frac{\pi}{T_1}}$$

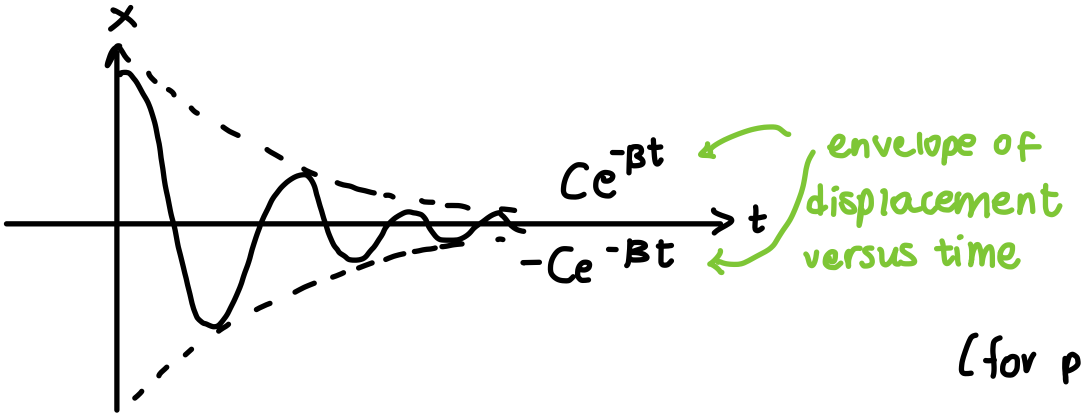
Note : the "angular frequency" of the damped oscillator is less than the frequency of tho oscillator in the absence of damping (i.e. $\omega_1 < \omega_0$).

Recall that

$\omega_1 = \sqrt{\omega_0^2 - \beta^2}$ if $\beta > 0$ $\quad \omega_1 < \omega_0$

The maximum amplitude of the motion of the damped oscillator decreases with time because of the factor $e^{-\beta t}$ (with $\beta > 0$). The envelope of the displacement versus time is given by

$$x_{env} = \pm C e^{-\beta t}$$



(for phase lag $\delta = 0$)

The ratio of the amplitudes of the oscillation at two successive maxima is

$$\frac{Ce^{-\beta T}}{Ce^{-\beta(T+2T_1)}} = \underbrace{e^{2\beta T_1}}_{} \quad \text{where} \quad 2T_1 = \frac{2\pi}{\omega_1} \Rightarrow T_1 = \frac{\pi}{\omega_1}.$$

↑ Called the DECREMENT of the motion

Critically damped motion

If $\beta^2 > \omega_0^2$ the system is prevented from undergoing oscillatory motion.

$$x(t) = e^{-\beta t}\left[Ae^{\underbrace{\sqrt{\beta^2-\omega_0^2}}_{\text{real}}\,t} + Be^{-\underbrace{\sqrt{\beta^2-\omega_0^2}}_{\text{real}}\,t}\right]$$

The case of critical damping occurs when $\boxed{\beta^2 = \omega_0^2}$

$$x(t) = e^{-\beta t}[A + B\underset{\uparrow}{t}]$$

since the roots are equal you need an extra t.

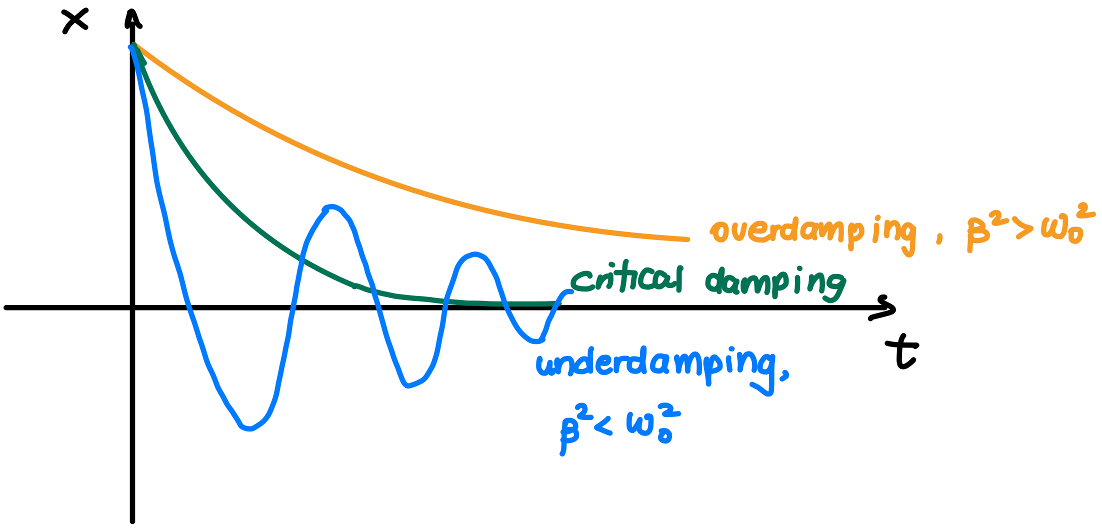
Overdamped motion

If the damping parameter β is larger than ω_0 ⇒ overdamping

Because $\beta^2 > \omega_0^2$, $x(t) = e^{-\beta t}[A\ e^{\omega_2 t} + Be^{-\omega_2 t}]$

where $\omega_2 = \sqrt{\beta^2 - \omega_0^2}$. Here ω_2 is <u>not</u> an angular frequency because the motion is <u>not</u> periodic

Overdamping results in a decrease of the amplitude to zero.



Example

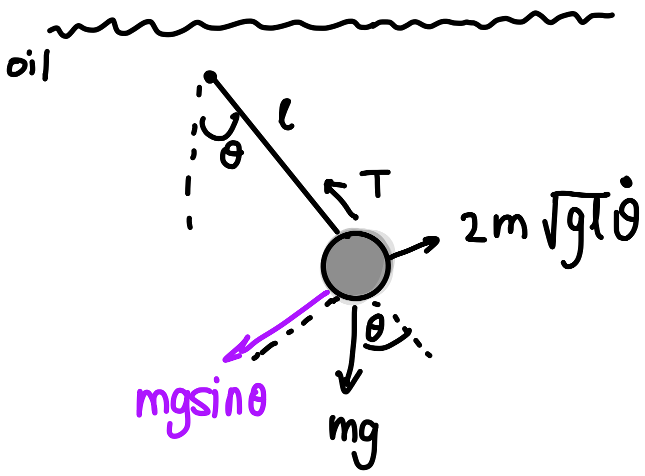
Consider a pendulum of length l and a mass m attached to the end, moving through oil with θ decreasing. The mass undergoes small oscillations, but the oil retards the mass' motion with a resistive force proportional to the speed, with $\boxed{F_{res} = 2m\sqrt{g/l}\; l\dot{\theta}}$ $\boxed{\dot{\theta} = \frac{d\theta}{dt}}$

The mass is initially pulled back at $t=0$ with $\theta = \alpha$ and $\dot{\theta} = 0$

Question Find the angular displacement θ and velocity $\dot{\theta}$ as a function of time.

Solution

$$\begin{aligned} \text{Force} &= ma \\ &= m(l\ddot{\theta}) \\ &= \text{restoring force} \\ &\quad + \text{resistive force} \end{aligned}$$



$$ml\ddot{\theta} = \underbrace{-mg\sin\theta}_{\text{restoring}} \underbrace{- 2m\sqrt{gl}\,\dot{\theta}}_{\text{resistive}}$$

For small oscillations $\sin\theta \simeq \theta$, so the equation becomes

$$\cancel{m}l\,\ddot{\theta} + \cancel{m}g\underset{\downarrow\,\theta}{\sin\theta} + 2\cancel{m}\sqrt{gl}\,\dot{\theta} = 0$$

$$\Rightarrow \quad \ddot{\theta} + \frac{g}{l}\theta + 2\sqrt{\frac{g}{l}}\,\dot{\theta} = 0$$

$$\Rightarrow \quad \boxed{\ddot{\theta} + 2\sqrt{\frac{g}{l}}\,\dot{\theta} + \frac{g}{l}\theta = 0}$$

Recall that for the damped oscillator the equation was given by

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0$$

and so if we compare the two, we see that

$$\beta = \sqrt{\frac{g}{l}} \quad \text{and} \quad \omega_0^2 = \frac{g}{l}$$

$$\Rightarrow \quad \beta^2 = \frac{g}{l}$$

which implies that $\boxed{\omega_0^2 = \beta^2}$ $\Rightarrow$ the pendulum is critically damped

We saw before that for a critically damped system the solution is

$$\theta(t) = (A + Bt)e^{-\beta t}$$

Using the initial conditions $\theta(0) = \alpha$ and $\dot{\theta}(0) = 0$ we can solve for A and B as follows

$$\theta(0) = \boxed{A = \alpha}$$

$$\dot{\theta} = Be^{-\beta t} + (A + Bt)(-\beta e^{-\beta t})$$

Using $\dot{\theta}(0)=0$ we have $0 = B + A(-\beta)$

$$\Rightarrow \quad 0 = B - \alpha\beta$$

$$\Rightarrow \quad \boxed{B = \alpha\beta}$$

Thus $\theta(t) = (\alpha + \alpha\beta t)e^{-\beta t}$ with $\beta = \sqrt{\frac{g}{l}}$

$$\Rightarrow \quad \theta(t) = \alpha\left(1+\sqrt{\frac{g}{l}}\,t\right)e^{-\sqrt{\frac{g}{l}}t}$$

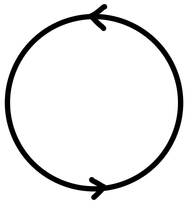
$$\dot{\theta}(t) = \cancel{\alpha\sqrt{\frac{g}{l}}\,e^{-\sqrt{\frac{g}{l}}t}} - \alpha\sqrt{\frac{g}{l}}\left(\cancel{1}+\sqrt{\frac{g}{l}}\,t\right)e^{-\sqrt{\frac{g}{l}}t}$$

$$= -\alpha\frac{g}{l}\,t\,e^{-\sqrt{\frac{g}{l}}t}$$

Lecture 4

(Chapter 4. Nonlinear dynamics and chaos by Strogatz)

Flows on the circle: $\boxed{\dot{\theta} = f(\theta)}$



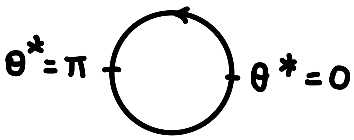
θ is a point on the circle

$\dot{\theta}$ is the velocity vector at that point.

By flowing in one direction, a particle can eventually return to its starting point. Thus periodic solutions become possible.

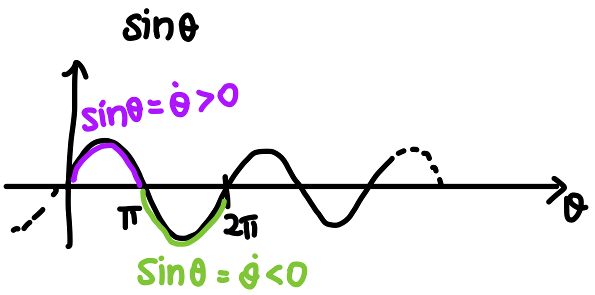
Example. Sketch the vector field on the circle corresponding to $\boxed{\dot{\theta} = \sin\theta}$.

Equilibrium points when $\dot{\theta}=0 \Rightarrow \sin\theta = 0 \Rightarrow \theta = 0, \pi$



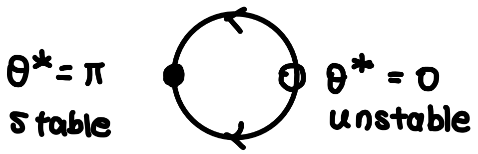
(counterclockwise as usual)

To find the stability of the equilibrium solutions we note that



This implies that for $0 \leq \theta \leq \pi$, $\dot{\theta} > 0 \Rightarrow$ θ increasing ⇒ moving counterclockwise

If $\pi \leq \theta \leq 2\pi$ then $\dot{\theta} < 0 \Rightarrow$ θ decreasing ⇒ moving clockwise



* * *

We need to assume that in $\dot{\theta} = f(\theta)$, $f(\theta)$ is a real-valued 2π-periodic function.

i.e. $\boxed{f(\theta + 2\pi) = f(\theta)}$ for all real θ. → for existence & uniqueness of solutions.

↓

This periodicity of f(θ) ensures that the velocity $\dot{\theta}$ is uniquely-defined at each point θ on the circle.

Uniform oscillator

A point on the circle is called an angle or a phase

The simplest oscillator is one in which the phase θ changes uniformly $\boxed{\dot{\theta} = \omega}$ for ω constant.

By integrating the equation we get that the solution is $\boxed{\theta(t) = \omega t + \theta_0}$

This is a uniform motion around the circle with an angular frequency ω.

Periodic with period $T = \frac{2\pi}{\omega}$.

Good way to obtain T:

$$\dot{\theta} = \frac{d\theta}{dt} = f(\theta) \Rightarrow \int_{\theta_0}^{\theta_0 + 2\pi} \frac{d\theta}{f(\theta)} = \int_0^T dt = T$$

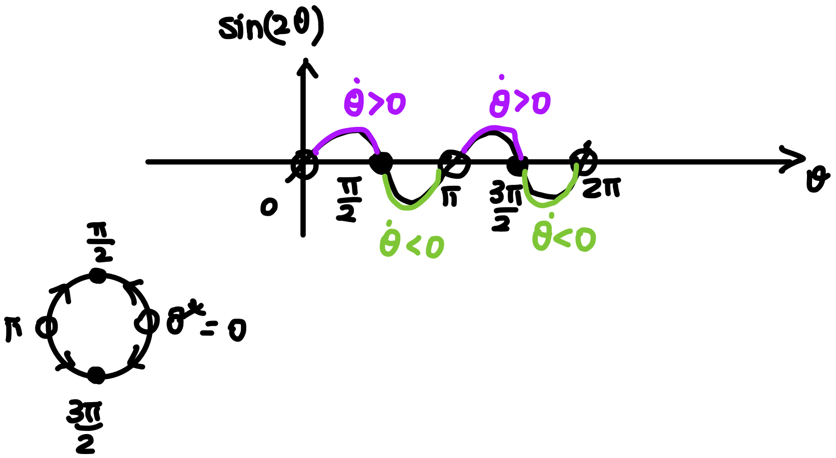
For $f(\theta) = \omega \Rightarrow T = [(\cancel{\theta_0} + 2\pi) - \cancel{\theta_0}]\frac{1}{\omega} = \frac{2\pi}{\omega}$

Example.

Find the equilibrium points of $\dot{\theta} = \sin(2\theta) = f(\theta)$ and determine their stability

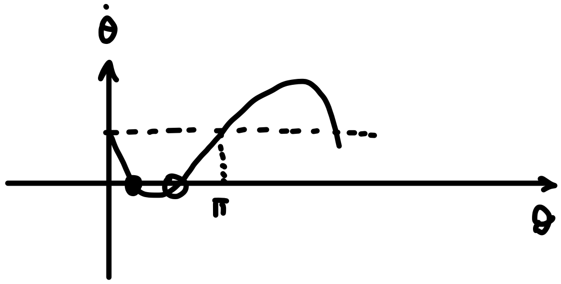
$\dot{\theta} = \sin(2\theta) = f(\theta)$

$\dot{\theta} = 0 \Rightarrow \sin(2\theta) = 0 \Rightarrow \theta = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$



Bifurcations

Consider $\dot{\theta} = \omega - a\sin\theta$, $\theta(0) = \theta_0$.



For equilibrium points:

$\theta = \omega - a\sin\theta = 0 \Rightarrow \sin\theta = \frac{\omega}{a}$

Lecture 5

3 cases:

$\frac{\omega}{a} < 1 \Rightarrow \sin\theta = \frac{\omega}{a} \Rightarrow \theta = \arcsin\left(\frac{\omega}{a}\right), \pi - \arcsin\left(\frac{\omega}{a}\right) \Rightarrow$ two eqm sol^{ns}

$\frac{\omega}{a} = 1 \Rightarrow \sin\theta = 1, \theta = \frac{\pi}{2} \Rightarrow$ one equilibrium solution

$\frac{\omega}{a} > 1 \Rightarrow$ no solutions to $\sin\theta = \frac{\omega}{a} > 1$ & so no eqm points

So if a is fixed and ω changed, note that we'll have no eqm points for $\omega > a$ and $\omega < -a$.

Do we really have two parameters?

We can do a change of variables to reduce this into a single-parameter problem.

$$\dot{\theta} = \omega - a\sin\theta$$

Divide by a throughout : $\frac{1}{a}\dot{\theta} = \frac{\omega}{a} - \sin\theta \Rightarrow \frac{1}{a}\frac{d\theta}{dt} = \frac{\omega}{a} - \sin\theta$

and let's define $\mu := \frac{\omega}{a}$ and $\tau = ta$. We'll get

$$d\tau = dt\, a$$

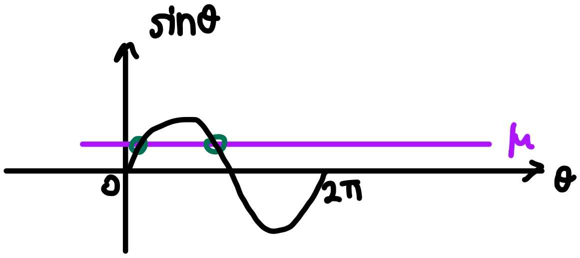
$$\frac{d}{d\tau} = \frac{1}{a}\frac{d}{dt}$$

Thus $\boxed{\frac{d\theta}{d\tau} = \mu - \sin\theta}$. Now we can use this one-control parameter eqn to analyze the system.

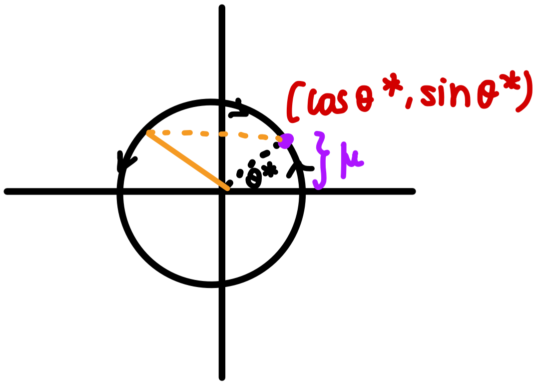
$$\mu - \sin\theta^* = 0 \Rightarrow \sin\theta^* = \mu$$

$\theta^* = \arcsin(\mu)$ solutions exist only for $|\mu| \leq 1$.

Let's now compute the stability of this problem:



Unit circle



The y-coord. is $\sin\theta = \mu$

$$\cos^2\theta + \sin^2\theta = 1 \Rightarrow \cos^2\theta = 1 - \sin^2\theta$$

$$\cos\theta = \pm\sqrt{1-\mu^2}$$

Recall $\begin{bmatrix} x = r\cos\theta \\ y = r\sin\theta \end{bmatrix}$. Unit circle $\Rightarrow r = 1$

$$\Rightarrow \begin{cases} x = \cos\theta \\ y = \sin\theta \end{cases}$$

Thus, we have 2 equilibria for $0 < \mu < 1$.

For the stability analysis we know that between $\theta^* = \arcsin(\mu)$ and $\theta^* = \pi - \arcsin(\mu)$, $\dot{\theta} = \mu - \sin\theta < 0$ and that between $\theta^* = 0$ and $\theta^* = \arcsin(\mu)$, $\dot{\theta} > 0$, which implies that $\theta^* = \arcsin(\mu)$ is stable.

However between $\theta^* = \pi - \arcsin(\mu)$ and $\theta = \pi$ we have $\dot{\theta} > 0$ which implies that $\theta^* = \pi - \arcsin(\mu)$ is unstable.

Question What's the total time to make one circle? e.g. generalized period.

$$\frac{d\theta}{dt} = f(\theta) \Rightarrow \int_{\theta_0}^{\theta_0 + 2\pi} \frac{d\theta'}{f(\theta')} = \int_0^T dt = T$$

Take $\theta_0 = 0$ $\Rightarrow$ $T = \int_0^{2\pi} \frac{d\theta}{\omega - a\sin\theta} \underset{\uparrow}{=} \int_{-\pi}^{\pi} \frac{d\theta}{\omega - a\sin\theta} = \int_{-\pi}^{\pi} \frac{d\theta}{\mu - \sin\theta}$

ok to shift to a different periodic interval
$\Rightarrow$ same answer

* * *

Fireflies Thousands of male fireflies flash on and off in unison.

They don't start out synchronized but the synchrony builds up gradually.

★ Fireflies influence each other ★ When one firefly sees the flash of another it slows down or speeds up so as to flash more closely in phase on the next cycle

MODEL

$\theta(t)$ = phase of the firefly's flashing rhythm

$\theta = 0$ corresponds to the instant when a flash is emitted

natural frequency

Without stimuli, the firefly goes through its cycle of frequency $\omega \Rightarrow \boxed{\dot{\theta} = \omega}$

Now suppose there's a periodic stimulus whose phase Θ satisfies $\boxed{\dot{\Theta} = \Omega}$ where $\Theta = 0$ corresponds to the flash of the stimulus.

Firefly's response to stimulus

If stimulus ahead in the cycle → firefly speeds up to synchronize

If it's flashing too early → firefly slows down

$$\boxed{\dot{\theta} = \omega - A\sin(\theta - \Theta)}\ , \text{ where } A>0$$

If θ is behind $\Theta \Rightarrow -\pi < \theta - \Theta < 0 \Rightarrow$ the firefly speeds up $(\dot{\theta} > \omega)$

If θ is ahead of $\Theta \Rightarrow 0 < \theta - \Theta < \pi \Rightarrow$ the firefly slows down $(\dot{\theta} < \omega)$

Model for 2 fireflies blinking

Each wants to sync with the other and each has different natural frequency

each is driven by the other

$$\left[\begin{array}{l} \dot{\theta}_1 = \omega_1 - a\sin(\theta_1 - \theta_2) \\ \dot{\theta}_2 = \omega_2 - a\underbrace{\sin(\theta_2 - \theta_1)}_{= -\sin(\theta_1-\theta_2)\text{ since sine is an odd function}} \end{array}\right.$$

↑ Same coupling strength

We define $\phi = \theta_1 - \theta_2 \Rightarrow \dot{\phi} = \dot{\theta}_1 - \dot{\theta}_2 = [\omega_1 - a\sin(\theta_1-\theta_2)] - [\omega_2 \underbrace{-a\sin(\theta_2-\theta_1)}_{+a\sin(\theta_1-\theta_2)}]$

$$= \omega_1 - \omega_2 - 2a\sin(\theta_1 - \theta_2)$$

$$= \omega_1 - \omega_2 - 2a\sin\phi$$

Model for 2 fireflies synchronizing flashes

Consider $\dot{\theta} = f(\theta)$, $\theta(0) = \theta_0$. This can model a periodic event, like a church bell ringing by assuming the ringing occurs when $\theta = 2n\pi$, $n \in \mathbb{Z}$

A bell that rings each hour would be modeled as a uniform oscillator

$$\boxed{\dot{\theta} = \omega}\ , \ \omega = 2\pi \text{ hour}^{-1}$$

$$T = \frac{2\pi}{\omega} = 1 \text{ hour}$$

Now we suppose that firefly 1 blinks when $\theta_1 = 2n\pi$ and also firefly 2 blinks when $\theta_2 = 2n\pi$. If measured individually, each has its own intrinsic frequency ω_1 & ω_2.

As above, we consider a coupled model

$$\dot{\theta}_1 = \omega_1 - a\sin(\theta_1 - \theta_2)$$
$$\dot{\theta}_2 = \omega_2 - a\sin(\theta_2 - \theta_1)$$

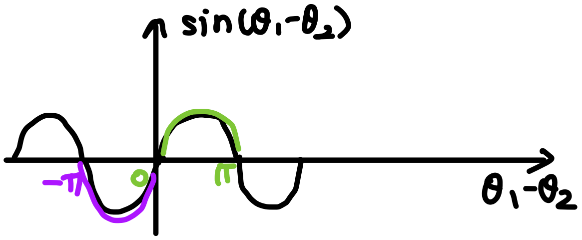
] think of them as speeds modified by phase Lag

① $\sin(\theta_1-\theta_2) > 0$ if $\theta_1 - \theta_2 \in (0, \pi)$

⇒ θ_1 leads

⇒ $\dot{\theta}_1 < \omega_1$, $\dot{\theta}_2 > \omega_2$

↑ θ_1 slows down



② $\sin(\theta_1-\theta_2) < 0$ if $\theta_1 - \theta_2 \in (-\pi, 0)$

⇒ θ_2 leads

⇒ $\dot{\theta}_1 > \omega_1$, $\dot{\theta}_2 < \omega_2$

↑

θ_1 speeds up to catch up w/ θ_2

Define phase difference

$$\phi = \theta_1 - \theta_2$$
$$\dot{\phi} = \dot{\theta}_1 - \dot{\theta}_2 = \underbrace{\omega_1 - \omega_2}_{\equiv \Delta\omega \text{ : demand } > 0} - 2a\sin\phi$$

$$\Rightarrow \boxed{\dot{\phi} = \Delta\omega - 2a\sin\phi}, \quad \Delta\omega = \omega_1 - \omega_2 \underbrace{\geq 0}_{\text{set}}$$

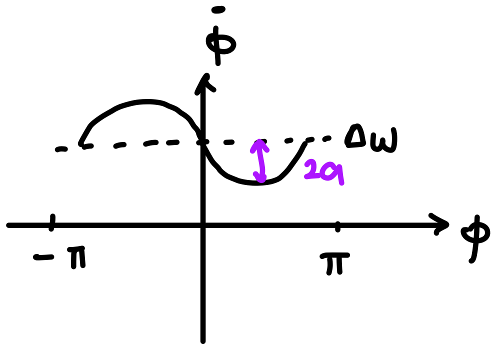
(choose fly w/ bigger ω as 1)

Note : Coupling strength a determine firefly's ability to modify its frequency

Consider different cases: $\dot{\phi} = \Delta w - 2a\sin\phi$

① $\Delta w > 2a \Rightarrow$ no equilibrium points

Blinking stays unsynced and out of phase

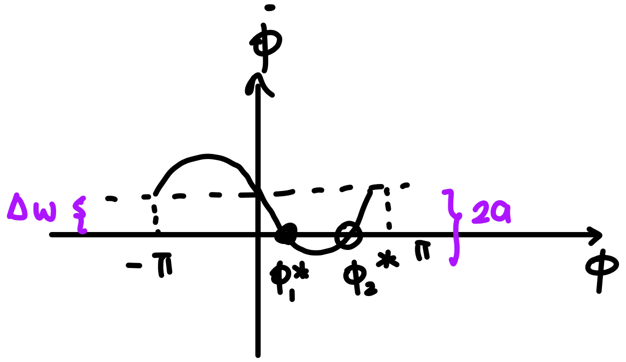


$\dot{\phi} = 0 \Rightarrow \Delta w - 2a\sin\phi = 0$

$\sin\phi = \frac{\Delta w}{2a} > 1$

$\Rightarrow$ no solutions

② $\Delta w < 2a \Rightarrow$ two equilibrium points, one stable.



For any initial conditions $\theta_1(0)$, $\theta_2(0)$, after sufficient time, the system will approach the equilibrium solutions and we'll have

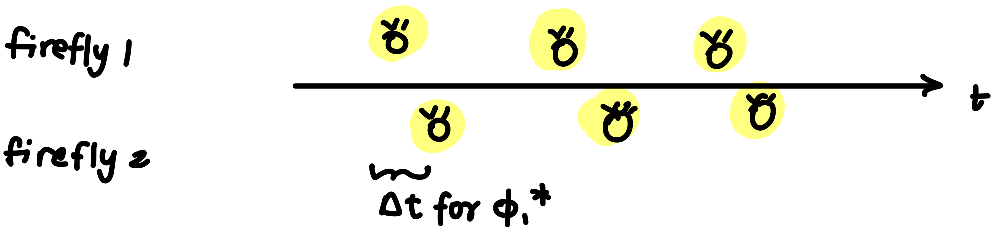
$$\theta_1(t) - \theta_2(t) = \phi_1^* > 0 \quad \text{const)}$$

Note $\phi(t) \rightarrow \phi_*$ and $\dot{\phi} = 0$

$\theta_1(t) - \theta_2(t) \rightarrow \phi_*$ & $\dot{\theta}_1 - \dot{\theta}_2 = 0$

Equilibrium point of $\phi \Rightarrow$ blinks at same frequency with phase lag ϕ_*

So they are in sync but slightly out of phase

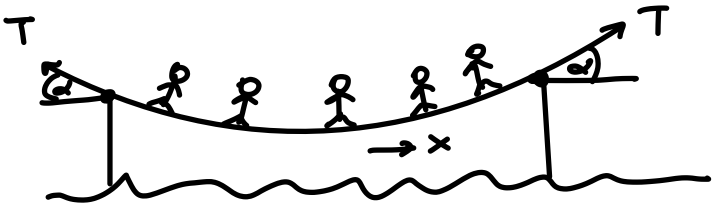


Lecture 6

(Chapter 14 in Tung's book)

Collapsing bridges We wish to model the oscillations of suspension bridges under forcing. (look up the collapse of the Tacoma Narrows Bridge as 1940)

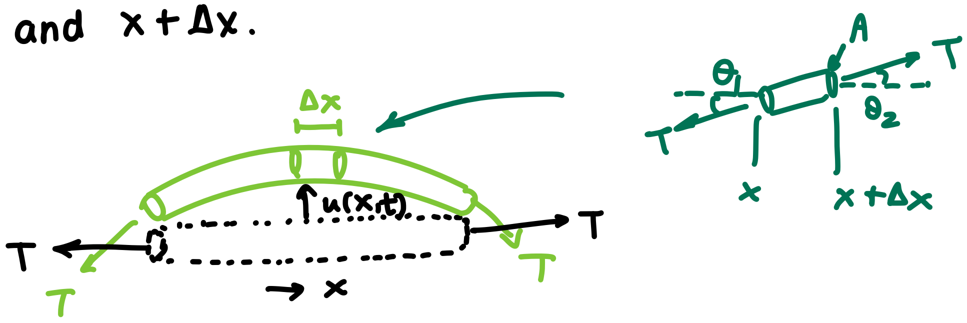
This is an example of resonance which happens when the frequency of forcing matches the natural frequeny of oscillation of the bridge.



When people march in unison over a bridge a vertical force $f(x,t)$ is exerted on the bridge that is periodic in time. w/ a period P determined by the time interval between steps.

We model the bridge as an elastic string of length L suspended only at $x=0$ and $x=L$

We consider the vertical displacement $u(x,t)$ of the string (bridge) from its equilibrium position, where x is the distance from the left suspension point and t is time. We consider a small section of the string between x and $x+\Delta x$.



We apply Newton's 2nd law of motion $\boxed{F = ma}$ to the vertical motion of this small section of the string.

Its mass is $\rho A \Delta x$ $\left(\rho = \frac{m}{V} \Rightarrow m = \rho V = \rho A \Delta x\right)$, where ρ is the density of the material of the string and A is its cross-sectional area. The acceleration in the vertical direction is $a = \frac{d^2u}{dt^2}$.

The force should be the vertical component of the tension, plus other forces such as gravity and air friction.

The net vertical component of tension is

$$T\sin\theta_2 - T\sin\theta_1 \approx T[\theta_2 - \theta_1]$$

assuming that θ_1, θ_2 are small

$$\simeq T\left[\frac{\partial u}{\partial x}(x+\Delta x, t) - \frac{\partial u}{\partial x}(x,t)\right]$$

↑

tension force per unit area

Putting everything together we have

$$\rho A \Delta x \frac{\partial^2 u}{\partial t^2} = TA\left[\frac{\partial u}{\partial x}(x+\Delta x, t) - \frac{\partial u}{\partial x}(x,t)\right] + \rho A \Delta x \cdot f$$

↑ $\frac{\partial u}{\partial x}$: stretch

↑ all additional force per unit mass

$(\div \rho A \Delta x)$

$$\Rightarrow \quad \frac{\partial^2 u}{\partial t^2} = \frac{T}{\rho}\frac{1}{\Delta x}\left[\frac{\partial u}{\partial x}(x+\Delta x, t) - \frac{\partial u}{\partial x}(x,t)\right] + f$$

and as $\Delta x \to 0$

$$\boxed{\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2} + f} \quad \text{where} \quad \boxed{c^2 \equiv \frac{T}{\rho}}$$

The tension along the bridge T is assumed to be uniform and is therefore equal to the force per unit area exerted on the subpension point $x=0$ or $x=L$.

Since the weight of the bridge is borne by these two suspension points, the vertical force exerted on each is half the weight of the bridge, and this is equal to the projection of T in the vertical direction

$$T\sin\alpha = \frac{1}{2}\frac{\overset{\text{mass}}{\boxed{(\rho L A)}}\,g}{A} = \frac{1}{2}\rho L g$$

where α = angle from horizontal to the tangent at the suspension point.

$$\Rightarrow \quad c^2 \equiv \frac{T}{\rho} = \cancel{\frac{1}{\rho}}\left(\frac{1}{2}\frac{\cancel{\rho} L g}{\sin\alpha}\right) = \frac{Lg}{2\sin\alpha}$$

Since the static weight of the bridge is balanced by tension, the forcing f represents unbalanced vertical acceleration due to the pedestrians.

The system we need to solve is $u(x,t)$ being the vertical displacement of the bridge wrt its equilibrium position

$$\left\{\begin{array}{lll} & \dfrac{\partial^2 u}{\partial t^2} = c^2\dfrac{\partial^2 u}{\partial x^2} + f(x,t), & 0<x<L,\ t>0 \\ \text{boundary conditions:} & u(0,t)=0, \quad u(L,t)=0, & t>0 \\ \text{initial conditions:} & u(x,0)=0, \ \dfrac{\partial u}{\partial t}(x,0)=0, & 0<x<L \end{array}\right\} \quad (*)$$

What's the form of the force function?

The simplest expression for the periodic force exerted by the pedestrians is

$$f(x,t) = a\sin(\omega_D t)\sin\left(\frac{\pi x}{L}\right), \quad \text{for } 0<x<L, \quad \omega_D = \frac{2\pi}{D}$$

Note that the form of the force function assumes that the pedestrians move in sync!

How do we solve the system (*)?

The solution will be a function of both space and time. We assume that we can write it in the separable form $\boxed{u(x,t) = X(x)T(t)}$ ← this must also satisfy the boundary & initial conditions

If we substitute this into the governing partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2} + f(x,t)$$

we obtain

$$XT'' = c^2X''T + a\sin(\omega_0 t)\sin\left(\frac{\pi x}{L}\right) \qquad (\dagger)$$

Next, we also assume that the form of $X(x)$ is known: $\boxed{X(x) \equiv \sin\left(\frac{\pi x}{L}\right)}$, $0<x<L$

Its second derivative w.r.t. space is $X''(x) = -\left(\frac{\pi}{L}\right)^2 \underbrace{\sin\left(\frac{\pi x}{L}\right)}_{X(x)} = -\left(\frac{\pi}{L}\right)^2 X(x)$

Substituting this into $(\dagger)$ we get

$$\cancel{X}T'' = c^2\left(-\left(\frac{\pi}{L}\right)^2\cancel{X}\right)T + a\sin(\omega_0 t)\cancel{X}$$

divide throughout by X

$$\boxed{T'' + \left(c\frac{\pi}{L}\right)^2 T = a\sin(\omega_0 t)} \qquad (\ddagger)$$

The "natural frequency" ω_1 of the bridge $\omega_1 = \frac{c\pi}{L}$. So we see that the natural frequency depends on L, which is the wavelength of the forcing structure.

Note that $T'' + \omega_1^2 T = a\sin(\omega_0 t)$ is an ODE rather than PDE.

Recall that when we covered the simple harmonic oscillator we derived the governing ODE: $\ddot{x} + \omega_0^2 x = 0$. Thus, $(\ddagger)$ is the ODE for the forced oscillator. So we saw that the natural frequency of the oscillator is related to the spatial structure of the oscillation.

Lecture 6

We now solve $(\ddagger)$. As we have done previously we find the characteristic eqn; for the homogeneous problem:

$$r^2 + \omega_1^2 = 0$$
$$r = \pm i\omega_1$$

$$\Rightarrow \quad T(t) = A\cos(\omega_1 t) + B\sin(\omega_1 t)$$

and for the particular solution we will try that $T(t) = c\sin(\omega_D t)$

So we begin by substituting $T(t) = c\sin(\omega_D t)$ into the ODE:

$$-c\omega_D^2 \sin(\omega_D t) + \omega_1^2 c\sin(\omega_D t) = a\sin(\omega_D t)$$

$$-c\omega_D^2 + \omega_1^2 c = a$$

$$c = \frac{a}{\omega_1^2 - \omega_D^2}$$

Thus, the particular solution is of the form $T(t) = \frac{a}{\omega_1^2 - \omega_D^2}\sin(\omega_D t)$.

This implies that the full general solution is

$$T(t) = \underbrace{A\cos(\omega_1 t) + B\sin(\omega_1 t)}_{\text{homogeneous sol}^n} + \underbrace{\frac{a}{\omega_1^2 - \omega_D^2}\sin(\omega_D t)}_{\text{particular sol}^n}$$

To find the solution for (*) we substitute the initial and boundary conditions

$u(0,t)=0, \quad u(L,t)=0$ ← boundary conditions

$u(x,0)=0, \frac{\partial u}{\partial t}(x,0)=0$ ← initial conditions

$$u(x,t) = \left[A\cos(\omega_1 t) + B\sin(\omega_1 t) + \frac{a}{\omega_1^2 - \omega_D^2}\sin(\omega_D t)\right]\sin\left(\frac{\pi x}{L}\right)$$

$u(0,t)=0 \Rightarrow$ identically zero

$u(L,t)=0 \Rightarrow$ identically zero.

$u(x,0)=0 \Rightarrow \quad A\sin\left(\frac{\pi x}{L}\right)=0 \Rightarrow \quad A=0$

Thus $u(x,t) = \left[B\sin(\omega_1 t) + \frac{a}{\omega_1^2 - \omega_D^2}\sin(\omega_D t)\right]\sin\left(\frac{\pi x}{L}\right)$

$$\frac{\partial u}{\partial t} = \left[B\omega_1\cos(\omega_1 t) + \frac{a\omega_D}{\omega_1^2 - \omega_D^2}\cos(\omega_D t)\right]\sin\left(\frac{\pi x}{L}\right)$$

$$\frac{\partial u}{\partial t}(x,0)=0 \Rightarrow \quad \left[B\omega_1 + \frac{a\omega_D}{\omega_1^2 - \omega_D^2}\right]\sin\left(\frac{\pi x}{L}\right) = 0$$

$$B = -\frac{a\omega_D}{\omega_1}\frac{1}{\omega_1^2-\omega_D^2}.$$

The solution is of the form:

$$u(x,t) = \left[-\frac{a\omega_D}{\omega_1}\frac{1}{\omega_1^2-\omega_D^2}\sin(\omega_1 t) + \frac{a}{\omega_1^2-\omega_D^2}\sin(\omega_D t)\right]\sin\left(\frac{\pi x}{L}\right)$$

$$= \frac{a}{\omega_1^2-\omega_D^2}\left[-\frac{\omega_D}{\omega_1}\sin(\underbrace{\omega_1}_{\text{natural frequency}} t) + \sin(\underbrace{\omega_D}_{\text{forced frequency}} t)\right]\sin\left(\frac{\pi x}{L}\right).$$

Resonance

The solution is valid for $\omega_1 \neq \omega_D$. Some special treatment is helpful when $\omega_D \to \omega_1$. We rewrite $\boxed{\omega_D = \omega_1 + \epsilon}$ and let $\epsilon \to 0$.

We rewrite $\frac{a}{\omega_1^2-\omega_D^2}\sin(\omega_D t)$ as

$$\frac{a\sin(\omega_1 t+\epsilon t)}{\omega_1^2-(\omega_1+\epsilon)^2} = \frac{a\sin(\omega_1 t+\epsilon t)}{\cancel{\omega_1^2}-\cancel{\omega_1^2}-2\epsilon\omega_1-\epsilon^2} = \frac{a\sin(\omega_1 t+\epsilon t)}{-2\epsilon\omega_1-\epsilon^2}$$

$$= \frac{a\sin(\omega_1 t)\cos(\epsilon t)+a\cos(\omega_1 t)\sin(\epsilon t)}{-2\epsilon\omega_1-\epsilon^2} = \frac{a\sin(\omega_1 t)\cos(\epsilon t)}{-2\epsilon\omega_1-\epsilon^2} + \frac{a\cos(\omega_1 t)\sin(\epsilon t)}{-2\epsilon\omega_1-\epsilon^2}$$

as $\epsilon \to 0$ $\cos(\epsilon t) \to 1$

$-2\epsilon\omega_1-\epsilon^2 \to -2\epsilon\omega_1$

$\frac{'0'}{'0'} \to$ L'Hôpital

$\to \frac{a\cos(\omega_1 t)\, t\cos(\epsilon t)}{-2\omega_1-2\epsilon}$

$\to \frac{a\cos(\omega_1 t)\, t}{-2\omega_1}$ as $\epsilon \to 0$

$$\to \frac{a\sin(\omega_1 t)}{-2\epsilon\omega_1} - \frac{a t\cos(\omega_1 t)}{2\omega_1} \quad \text{as } \epsilon \to 0$$

$$u(x,t) = a\left[-\frac{t\cos(\omega_1 t)}{2\omega_1} + \frac{\sin(\omega_1 t)}{2\omega_1^2}\right]\sin\left(\frac{\pi x}{L}\right)$$

→ Grows linearly in time

⇒ collapse of bridge

The fundamental frequency ω_1 is given by $\omega_1 = \frac{c\pi}{L}$. Recall that $c^2 = \frac{Lg}{2\sin\alpha}$ and so $\omega_1 = \sqrt{\frac{Lg}{2\sin\alpha}}\,\frac{\pi}{L} \Rightarrow \omega_1 = \pi\sqrt{\frac{g}{2L\sin\alpha}}$

Thus, the natural period P_1 is given by $P_1 = \frac{2\pi}{\omega_1} = \frac{2\cancel{\pi}}{\cancel{\pi}\sqrt{\frac{g}{2L\sin\alpha}}} = \sqrt{\frac{8L\sin\alpha}{g}}$

So if the bridge is $L \sim 10\,m$ long and the bridge deck is nearly horizontal $\alpha \sim 10°$ then

$$P_1 = \sqrt{\frac{8(10)\sin(10\pi/180)}{g}} = 1.1906 \text{ seconds}$$

This is close to the probable forcing period P, and resonance is likely. Note that there is no need for an exact match of the two frequencies to get an enhanced response.

* * *

Discovery of dynamical systems using regression

(modified notes of Karthik Duraisamy)

Consider nonlinear systems and try to discover their structure, purely based on observations of the system. What we are ultimately after is not just a model that explains the data but rather the governing equations themselves, so that we can confidently make predictions far from the training data.

Setup Start with the dynamical system

$$\boxed{\vec{x}^{\,n+1} = \vec{f}(\vec{x}^{\,n}) \quad ; \quad \vec{x}(0) = \vec{x}^{\,0} \; ; \; \vec{x} \in \mathbb{R}^N .}$$

We are asking the following question:

> If we just have some data (either the state $\vec{x}$ or some observable of the state $\vec{g}(\vec{x})$ at some time instances), can we recover the dynamical system above?

Note that we're not interested in reconstructing a solution that we've already seen nor are we just interested in interpolation. We want to make predictions far away from the data. To do this, we need to extract the functional form of $\vec{F}$ from data.

__The key idea__. Consider a nonlinear dynamical system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu x_1 \\ \lambda(x_2 - x_1^2) \end{pmatrix}$$

Define a set of features $\vec{\Psi}(\vec{x}) = \begin{pmatrix} x_1 \\ x_2 \\ x_1^2 \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix}$

Then a linear system of equations can be written for the evolution of $\vec{\Psi}(\vec{x})$.

$$\frac{d}{dt}\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 2\mu \end{pmatrix}\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix}$$

- $\frac{d\psi_1}{dt} = \frac{dx_1}{dt} = \mu x_1 = \mu \psi_1$
- $\frac{d\psi_2}{dt} = \frac{dx_2}{dt} = \lambda x_2 - \lambda x_1^2 = \lambda \psi_2 - \lambda \psi_3$
- $\frac{d\psi_3}{dt} = \frac{d}{dt} x_1^2 = 2x_1 \frac{dx_1}{dt} = 2x_1(\mu x_1)$

$$= 2\mu x_1^2 = 2\mu \psi_3$$

So, what have we __gained__ here?

We've taken a nonlinear ODE system for $\vec{x}$ and transformed it to a linear ODE system for $\vec{\Psi}$, without any loss of information or accuracy!

__Penalty__. We have increased the dimension of this system

This opens the door to tools such as linear regression to extract the underlying system of equations.

__Lecture 7__

Nonlinear approximations by transforming to feature space.

Assume we are given M data points

$$\vec{X} = (\vec{x}_1 \;\; \vec{x}_2 \;\; \ldots \;\; \vec{x}_m) \quad \text{input}$$

and output $\vec{Y} = (\vec{y}_1 \;\; \vec{y}_2 \cdots \vec{y}_m)$ where $\vec{y} = \vec{F}(\vec{x})$

__Note__ $\vec{x}_j$ and $\vec{x}_{j+1}$ do __not__ have to be in sequence.

To continue we need some basis functions which we will refer to as features. We define a feature vector $\vec{\psi}(\vec{x}) \in \mathbb{R}^p$

$$\vec{\psi}(\vec{x}) = \begin{pmatrix} \psi_1(\vec{x}) \\ \psi_2(\vec{x}) \\ \vdots \\ \psi_p(\vec{x}) \end{pmatrix}$$

dimensions are p×1

← these features $\psi_k(\vec{x})$ can be for example polynomials

Define a features-to-state matrix $\vec{C}$ in the following way.

$$\underbrace{\vec{x}}_{N\times 1} = \underset{\overset{\uparrow}{N\times P}}{\vec{C}}\, \underbrace{\vec{\psi}(\vec{x})}_{P\times 1}$$

In many situations, $\vec{C}$ could be trivial as it makes sense to have $\vec{x}$ as one of the features. To be formal, defining $\overset{P\times m}{\vec{\Psi}_x} = [\vec{\psi}(\vec{x}_1)\ \vec{\psi}(\vec{x}_2)\ \vec{\psi}(\vec{x}_3) \dots \vec{\psi}(\vec{x}_M)]$ we can obtain $\vec{C}$ via

$$\boxed{\underset{N\times P}{\vec{C}} = \overset{N\times m}{\vec{X}}\, \underset{m\times P}{\vec{\Psi}_x^{+}}}$$

but $\vec{\Psi}_x^{\underset{\text{inverse}}{+}}$ has dimensions m×P

Similarly, define $\vec{\Psi}_y = [\vec{\psi}(\vec{y}_1)\ \vec{\psi}(\vec{y}_2) \dots \vec{\psi}(\vec{y}_m)]$

Now we know that in the state space the system goes from one time step to the next in a nonlinear fashion. However, we could look for a linear update in feature space

$$\vec{\Psi}_y \approx \vec{K}\vec{\Psi}_x$$

and determine $\vec{K}$ by a least squares minimization over the data

$$\vec{K} = \vec{\Psi}_y \vec{\Psi}_x^{+}$$

Then we have $\vec{X} = \vec{C}\vec{\Psi}_x$ and $\vec{Y} = \vec{C}\vec{\Psi}_y = \vec{C}\vec{K}\vec{\Psi}_x$

Once $\vec{K}$ and $\vec{C}$ have been obtained, we can use then for any $\hat{\vec{x}}$.

$$\underset{N\times 1}{\vec{\hat{x}}^{n+1}} = [\underset{N\times p}{\vec{C}}\overset{}{\underset{p\times p}{\vec{K}}}]\underbrace{\vec{\Psi}(\vec{\hat{x}}^{n})}_{p\times 1}$$

Note that $\vec{C}$ and $\vec{K}$ are pre-computed matrices

SHOW MATLAB CODE

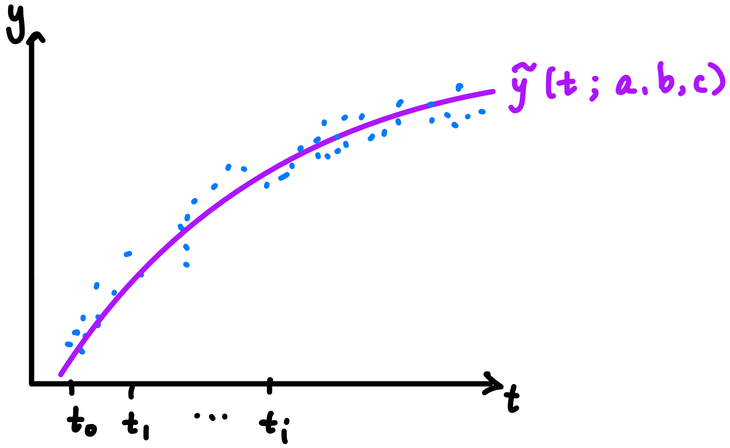
(Modified notes of Shafer Smith)

Parameter estimation with Gauss-Newton

Given data $y(t_i) = y_i$, $i = 1, 2, \dots, N$ and model $\tilde{y}(t; \theta_1, \dots, \theta_j)$ with $j = 1, \dots, M$

Find optimal $\boxed{\text{parameters}}$ $\theta_1, \dots, \theta_j$.

Example Suppose we have $\boxed{\text{data } y_i = y(t_i)}$ $i = 1, \dots, N$



and we believe the model is $\tilde{y}(t; \underbrace{a, b, c}) = at + b\cdot \ln(t+c)$

parameters

$\underline{\theta} = (\theta_1, \dots, \theta_M)$

$\underline{\theta} = [a, b, c]$

$M = 3$

NB any useful model will have $M \ll N$!

① Define a "cost function": $C(\vec{\theta})$

$$C(\vec{\theta}) = \sum_{i=1}^{N} [y_i - \tilde{y}(t_i, \vec{\theta})]^2 > 0 \text{ unless model fits the data exactly}$$

② Find its minimum wrt parameters θ_j, $j=1,\dots,M$

$$\frac{\partial C}{\partial \theta_j} = 0 \quad ; \quad j \text{ equations}$$

$$\frac{\partial C}{\partial \theta_j} = 2\sum_{i=1}^{N} [y_i - \tilde{y}(t_i;\vec{\theta})]\left(-\frac{\partial \tilde{y}}{\partial \theta_j}\right) = 0$$

This implies that $\boxed{\sum_{i=1}^{N} [y_i - \tilde{y}(t_i;\vec{\theta})] \frac{\partial \tilde{y}}{\partial \theta_j} = 0}$ for $j=1,\dots,M$

Solve these M equations for $\theta_1,\dots,\theta_M$

<u>Example</u> Model for data plotted above:

$$\tilde{y}(t,\vec{\theta}) = \theta_1 t + \theta_2 \ln(t+\theta_3)$$

Derivatives wrt parameters:

$$\frac{\partial \tilde{y}}{\partial \theta_1} = t$$

$$\frac{\partial \tilde{y}}{\partial \theta_2} = \ln(t+\theta_3)$$

$$\frac{\partial \tilde{y}}{\partial \theta_3} = \frac{\theta_2}{t+\theta_3}$$

Now let's consider a simpler case to see how to proceed

... Consider a special case with model that depends linearly on $\underline{\theta}$:

$$\tilde{y}(t,\vec{\theta}) = \theta_1 f_1(t) + \dots + \theta_M f_M(t) = \vec{\theta}\cdot\vec{f}(t)$$

$$\left[\text{e.g.} \quad \tilde{y}(t;a,b,c) = a\underset{\substack{\shortparallel \\ f_1(t)}}{t} + b\underset{\substack{\shortparallel \\ f_2(t)}}{t^2} + c\underset{\substack{\shortparallel \\ f_3(t)}}{\ln(t)} \right]$$

$$\Rightarrow \quad C(\vec{\theta}) = \sum_{i=1}^{N} \left[y_i - \vec{f}(t)\cdot\vec{\theta}\right]^2 \qquad \text{"cost-function"}$$

$$\Rightarrow \quad \frac{\partial C}{\partial \theta_j} = 0 \longrightarrow \sum_{i=1}^{N} \left[y_i - \vec{f}(t_i)\cdot\vec{\theta}\right] \underset{\overset{\uparrow}{\frac{\partial \tilde{y}}{\partial \theta_j}}}{f_j(t_i)} = 0$$

Lecture 8

Definition: $A_{ij} = f_j(t_i)$, $A = \underbrace{\left[\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_M\right]}_{M \text{ columns}} \Big\}\, N \text{ rows}$

$$\text{with } \vec{f}_j = \begin{bmatrix} f_j(t_1) \\ \vdots \\ f_j(t_N) \end{bmatrix}$$

$$\Rightarrow \quad \sum_{i=1}^{N} \left[y_i - \vec{f}(t_i)\cdot\vec{\theta}\right] f_j(t_i) = 0$$

OR

$$\sum_{i=1}^{N} \left[y_i A_{ij} - A_{ij} \sum_{s=1}^{M} \underbrace{f_s(t_i)}_{A_{is}} \theta_s\right] = 0$$

OR

$$y_i A_{ij} - A_{ij} \underbrace{A_{is}\theta_s}_{A\theta} = 0$$

← "Einstein notation" sum over repeated indices in products

Definition $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$ and $\vec{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_M \end{bmatrix}$

$$\Rightarrow \quad \boxed{A^T\vec{y} - A^TA\vec{\theta} = \vec{0}} \qquad \text{Least squares}$$

$$\Rightarrow \quad \vec{\theta} = (A^TA)^{-1}(A^T\vec{y})$$

Show that $\underbrace{A^TA}_{\text{size } m\times m}$ is symmetric

$\Rightarrow (A^TA)^{-1}$ exists

$A\vec{\theta} = \vec{y}$ →

$$N\left\{\underbrace{\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}}_{m}\right. \underbrace{\begin{bmatrix}\vec{\theta}\end{bmatrix}}_{1}\Big\}M = \left.\underbrace{\begin{bmatrix} \\ \vec{y} \\ \\ \end{bmatrix}}_{1}\right\}N$$

overdetermined system

N equations with m<<N unknowns

$N \times m$ $m \times 1$ $N \times 1$

Example

$\boxed{m=3}$ $\theta_1 \vec{f}_1 + \theta_2 \vec{f}_2 + \theta_3 \vec{f}_3 = \vec{y}$

unlikely that $\vec{y}$ lies in span of column space of A

Least squares Project $\vec{y}$ onto column space of A :

$$\underbrace{\underset{M\times N}{A^T}\,\underset{N\times M}{A}}_{M\times M}\,\underset{M\times 1}{\vec{\theta}} = \underset{M\times N}{A^T}\,\underset{N\times 1}{\vec{y}}$$

overall : $M \times 1$ both sides

In general the model will be nonlinear in $\vec{\theta}$.

→ Linearize model : first order Taylor expansion

→ $\tilde{y}(t, \vec{\theta}^{(1)}) \approx \tilde{y}(t, \vec{\theta}^{(0)}) + \sum_{j=1}^{M} \frac{\partial \tilde{y}}{\partial \theta_j} \cdot (t, \vec{\theta}^{(0)}) \cdot (\theta_j^{(1)} - \theta_j^{(0)})$

iterate

GUESS

Cost minimization functions :

$$\sum_{i=1}^{N} \Big[\underbrace{y_i - \underbrace{\tilde{y}(t_i, \theta^{(0)})}_{\tilde{y}_i^{(0)}}}_{\Delta y_i^{(0)}} - \sum_{s=1}^{M} \frac{\partial \tilde{y}_i^{(0)}}{\partial \theta_s} \cdot \underbrace{(\theta_s^{(1)} - \theta_s^{(0)})}_{\Delta\theta_s^{(1)}}\Big) \Big] \cdot \frac{\partial \tilde{y}_i^{(0)}}{\partial \theta_j} = 0$$

$\Delta\theta_s^{(1)}$: unknown

SOLVE FOR $\theta_s^{(1)}$ $s=1,\cdots,M$

Definition $J_{ij}^{(0)} := \frac{\partial \check{y}}{\partial \theta_j}(t_i, \vec{\theta}^{(0)})$

$$\rightarrow \left[\Delta y_i^{(0)} - J_{is}^{(0)} \Delta\theta_s^{(1)}\right] J_{ij}^{(0)} = 0$$

$$\Rightarrow \boxed{J^T J \,\Delta\vec{\theta}^{(1)} = J^T \Delta\vec{y}}$$

Normal equations

- Solve for $\Delta\vec{\theta}^{(1)}$ $\Rightarrow$ $\vec{\theta}^{(1)} = \Delta\vec{\theta}^{(1)} + \vec{\theta}^{(0)}$
  m values
  improved estimate

- Start again with guess $\vec{\theta}^{(1)} \xrightarrow[\text{solve normal equations}]{} \vec{\theta}^{(2)}$

- Repeat to $\vec{\theta}^{(k)}$

Stop when $\|\Delta\vec{\theta}^{(k)}\| = \|\vec{\theta}^{(k)} - \vec{\theta}^{(k-1)}\| < \text{Tol}$ ← some tolerance set in the code

i.e. when $\vec{\theta}^{(k)}$ is not changing "much" from $\vec{\theta}^{(k-1)}$

NB In a code we'd also define a maximum number of iterations k to prevent an infinite loop if $\|\Delta\theta^{(k)}\|$ never gets below the tolerance value we set.

This is an application of Newton's method for finding roots.

Angular momentum of a particle

(Kleppner-Kolenkow book)

Angular momentum $\vec{L}$ of a particle that has momentum $\vec{p} = m\vec{v}$ and is at position $\vec{r}$ w.r.t. a given origin:

$$\boxed{\vec{L} = \vec{r} \times \vec{p}} \quad \text{where } |\vec{L}| = L = rp\sin\alpha$$

angle between $\vec{r}$ and $\vec{p}$

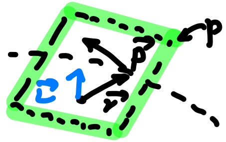
by definition of cross product

$$\vec{L} = |\vec{r}||\vec{p}|\sin(\alpha)\,\hat{k} = rp\sin(\alpha)\,\hat{k} = L_z\hat{k}$$

Remarks

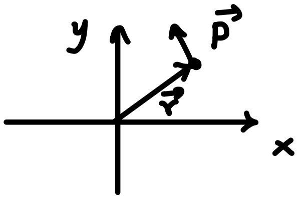
- $\vec{p}$ is independent of the coordinate system but $\vec{L}$ is not.
- $\vec{L}$ is perpendicular to the plane of motion

e.g. if $\vec{r}$ and $\vec{p}$ lie in the x-y plane, $\vec{L}$ lies along the z-direction

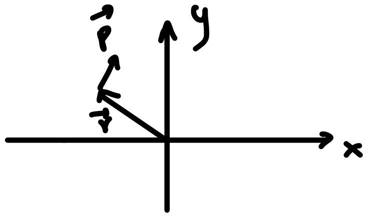


plane of motion

The right-hand rule determines if it is in the positive or negative z-directions: Point your fingers (right hand) along $\vec{r}$ and orient your hand so that you bend your fingers toward $\vec{p}$; your thumb then points in the direction of L.

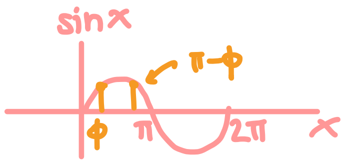


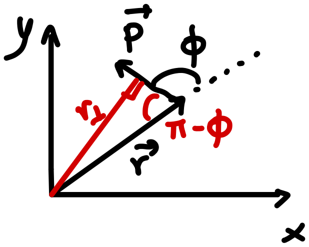
$$\boxed{L_z > 0}$$



$$\boxed{L_z < 0}$$

Geometrical understanding





Decompose $\vec{r}$ into $r_\perp$ that is perpendicular to the trajectory and $r_\parallel$ that is parallel

$$r_\perp = r\sin(\pi-\phi) = r\sin\phi$$

$\sin(\pi-\phi) = \frac{r_\perp}{r}$

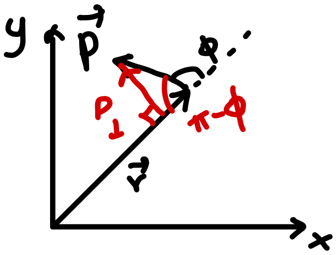
$\Rightarrow r_\perp = r\sin(\pi-\phi)$

Recall

$\vec{L} = \vec{r}\times\vec{p} = L_z\hat{k}$

where $L_z = rp\sin\phi$

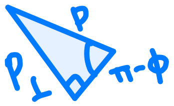
$$\Rightarrow L_z = \underbrace{r p\sin\phi}_{r_\perp} = r_\perp p$$



Decompose $\vec{p}$ into $p_\perp$ that is perpendicular to $\vec{r}$ and $p_\parallel$ that is parallel

$$p_\perp = p\sin(\pi-\phi) = p\sin\phi$$

$$\Rightarrow L_z = r\underbrace{p\sin\phi}_{p_\perp} = rp_\perp$$



$\sin(\pi-\phi) = \frac{p_\perp}{p}$

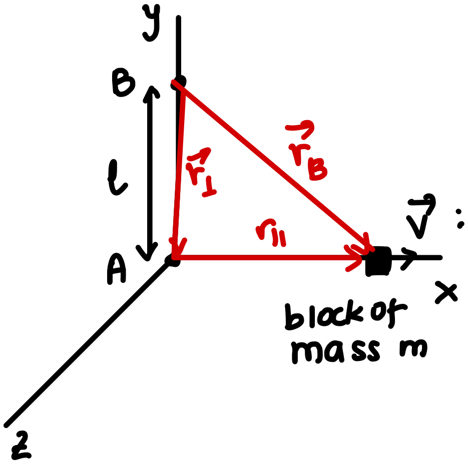
$p_\perp = p\sin(\pi-\phi)$

$= p\sin(\phi)$

Algebraically: $\vec{r} = (x, y, 0)$, $\vec{p} = (mv_x, mv_y, 0)$

$$\Rightarrow \vec{L} = \vec{r}\times\vec{p} = m\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ x & y & 0\\ v_x & v_y & 0\end{vmatrix} = m(xv_y - yv_x)\hat{k}$$

Example Angular momentum of a sliding block



$\vec{v}$: sliding freely in the x-direction ($\vec{v} = v\hat{i}$)

$$\vec{L}_A = m\underbrace{\vec{r}_A\times\vec{v}}_{\vec{r}\times\vec{p}\text{ where }\vec{p}=m\vec{v}} = 0$$

$$\vec{L}_B = m\vec{r}_B\times\vec{v} = m(\vec{r}_\parallel + \vec{r}_\perp)\times\vec{v} = mlv\hat{k}$$

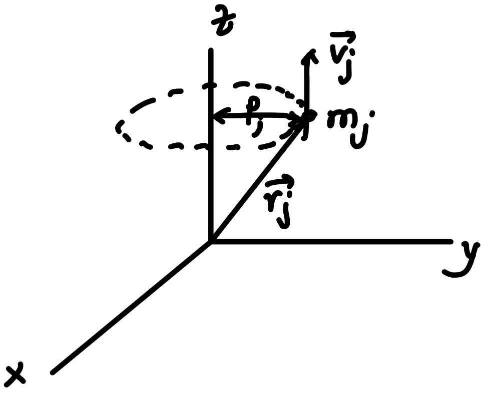
or $$\vec{L}_B = m\begin{vmatrix}\hat{i} & \hat{j} & \hat{k}\\ x & -l & 0\\ v & 0 & 0\end{vmatrix} = mlv\hat{k}$$

Fixed axis rotation

The direction of the axis of rotation is always along the same line, e.g. a car wheel attached to an axle undergoes fixed axis rotation as long as the car drives straight.

- When a rigid body rotates around an axis, every particle in the body remains at a fixed distance from the axis
- A coordinate system with its origin on the axis, $|\vec{r}| = \text{const}$ for every particle
  → $\vec{r}$ changes while $|\vec{r}|$ remains const: velocity is perpendicular to $\vec{r}$.

Consider a body rotating around the z-axis:



ρ_j : perpendicular distance to the axis of rotation from particle m_j

$$\rho_j = \sqrt{x_j^2 + y_j^2}$$

$$|\vec{v}_j| = \rho_j \cdot \underset{\uparrow}{\omega_j}$$

rate of rotation (angular speed)

Angular momentum of th j^{th} particle:

$$\vec{L}_j = \vec{r}_j \times m_j \cdot \vec{v}_j$$

→ not exactly in the z-direction!

Our focus: the component of angular momentum along the axis of rotation (z here)

$$\Rightarrow \quad L_{j,z} = \rho_j m_j v_j = m_j \rho_j^2 \omega_j$$

sum over all particles of the body

$$\text{For the whole body} \quad L_z = \sum_j L_{j,z} = \overset{\downarrow}{\sum_j} m_j \rho_j^2 \omega_j = \sum_j m_j \rho_j^2 \underset{\uparrow}{\omega}$$

ω is constant
(rigid body)

Torque $\boxed{\vec\tau = \vec r \times \vec F}$ torque due to force $\vec F$ that acts on a particle at position $\vec r$

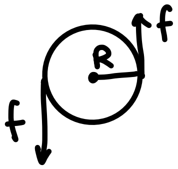
From above $\vec\tau = |\vec r||\vec F|\sin(\phi)\,\hat k$
$= r\sin\phi\,|\vec F|\,\hat k$
$= r_\perp |\vec F|\,\hat k$ and similarly $|\vec r||\vec F_\perp|\,\hat k$

$|\vec\tau| = |\vec r_\perp|\cdot|\vec F| = |\vec r|\cdot|\vec F_\perp|$

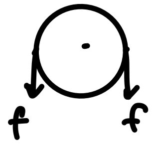
Also, $\vec\tau = \begin{vmatrix} \hat i & \hat j & \hat k \\ x & y & z \\ F_x & F_y & F_z \end{vmatrix}$ ⤳ torque depends on the origin we choose but force does not

$\vec\tau$ and $\vec F$ are always mutually perpendicular

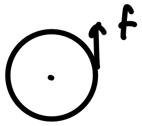
- Force and torque are inherently different quantities



$F=0$
$\tau = 2Rf$



$F=2f$
$\tau=0$



$F=f$
$\tau = Rf$

three different cases of τ, F combinations (τ is evaluated around the center of the disk)

Torque due to gravity

For a uniform gravitational field: $\vec\tau = \vec R \times \vec W$ ← weight
↑ vector to the center of mass

Proof. $\vec\tau_j = \vec r_j \times m_j \vec g = m_j \vec r_j \times \vec g$

$$\Rightarrow \vec\tau = \sum_j \vec\tau_j = \Big(\sum_j m_j \vec r_j\Big)\times\vec g \Rightarrow \vec\tau = \vec R \times M\vec g$$

Lecture 9

Torque and angular momentum

$\vec L = \vec r \times \vec p$

$$\Rightarrow \frac{d\vec L}{dt} = \frac{d\vec r}{dt}\times\vec p + \vec r \times \frac{d\vec p}{dt} = \underbrace{\vec v \times \vec p}_{=0} + \vec r \times \boxed{\frac{d\vec p}{dt}}$$

$\Rightarrow \vec F = \frac{d\vec p}{dt}$

by Newton's 2nd law

$\vec p = m\vec v$

Thus $\frac{d\vec{L}}{dt} = \vec{r} \times \vec{F} = \vec{\tau}$

Altogether

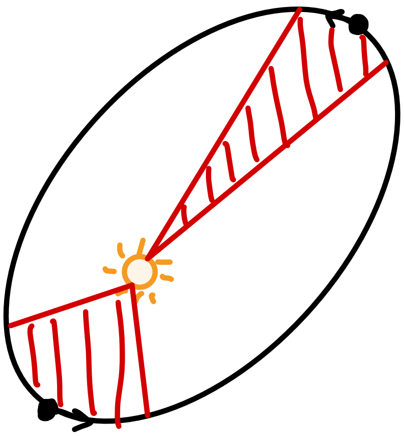
$$\boxed{\begin{array}{l} \frac{d\vec{L}}{dt} = \vec{\tau} \\ \frac{d\vec{p}}{dt} = \vec{F} \end{array}}$$

If $\vec{\tau} = 0$ then $\frac{d\vec{L}}{dt} = 0 \Rightarrow \vec{L}$ is constant and angular momentum is conserved.

Law of equal areas (Kepler's second law)

Explanation: Earth is moving under a central force (gravity, but can be extended to any central force)

$\vec{F}(\vec{r}) = f(r)\hat{r}$ ← unit vector in the radial direction



the area swept by the Earth for a given time is constant.

(equal areas in equal time)

- shorter radius
- higher speed

$\vec{\tau} = \vec{r} \times \vec{F} = \vec{r} \times f(r)\hat{r} = 0$

around Sun

$\Rightarrow$ the angular momentum is conserved

$\vec{L}$ is therefore constant in both magnitude and direction

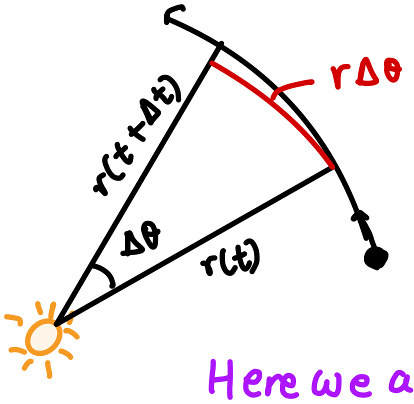
$\Rightarrow$ motion is confined to a plane!

For small Δθ, the area swept by the Earth can be approximated as

$$\Delta A \approx \frac{1}{2}(r(t+\Delta t))\cdot(r\Delta\theta)$$

$$= \frac{1}{2}(r+\Delta r)\cdot(r\Delta\theta)$$

$$= \frac{1}{2}r^2\Delta\theta + \frac{1}{2}r\Delta r\Delta\theta$$



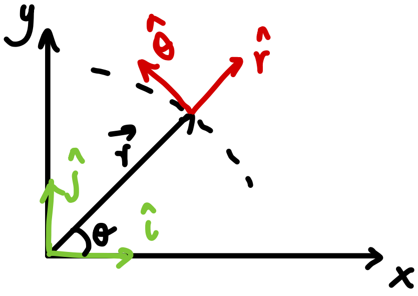
Here we assume that for very small Δθ, there is no difference betn an elliptical sector and a circular sector

The rate at which area is swept is

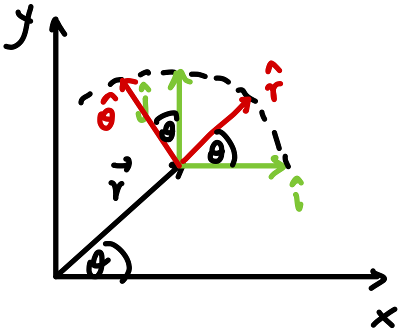
$$\frac{dA}{dt} = \lim_{\Delta t\to 0}\frac{\Delta A}{\Delta t} = \lim_{\Delta t\to 0}\left[\frac{1}{2}r^2\frac{\Delta\theta}{\Delta t} + \frac{1}{2}r\frac{\cancelto{0}{\Delta r}\Delta\theta}{\Delta t}\right]$$

$$\rightarrow \boxed{\frac{dA}{dt} = \frac{1}{2}r^2\dot{\theta}}$$

A short detour to polar coordinates...



Fundamental difference:
the directions of $\hat r$ and $\hat\theta$ vary with position, whereas $\hat i$ and $\hat j$ have fixed directions



$$\boxed{\begin{aligned}\hat r &= \hat i\cos\theta + \hat j\sin\theta\\ \hat\theta &= -\hat i\sin\theta + \hat j\cos\theta\end{aligned}}$$

So we can write $\vec r = r\cos\theta\,\hat i + r\sin\theta\,\hat j$

$$= r(\underbrace{\cos\theta\,\hat i + \sin\theta\,\hat j}_{\hat r})$$

$$= r\hat r$$

$$\boxed{\vec r = r\hat r}$$

Velocity in polar coordinates:

$$\frac{d\vec{r}}{dt} = \frac{dr}{dt}\hat{r} + r\frac{d\hat{r}}{dt} \quad = \dot\theta\hat\theta$$

$$\Rightarrow \boxed{\frac{d\vec{r}}{dt} = \dot{r}\hat{r} + r\dot\theta\hat\theta}$$

(velocity)

$$\frac{d\hat{r}}{dt} = \frac{d}{dt}(\cos\theta\,\hat{i} + \sin\theta\,\hat{j})$$

$$= -\sin\theta\underset{\sim}{\frac{d\theta}{dt}}\hat{i} + \cos\theta\underset{\sim}{\frac{d\theta}{dt}}\hat{j}$$

$$= \frac{d\theta}{dt}(\underbrace{-\sin\theta\,\hat{i} + \cos\theta\,\hat{j}}_{\hat\theta})$$

$$= \dot\theta\hat\theta$$

Finally, we also compute the acceleration which is the rate of change of velocity

$$\vec{a} : \frac{d\vec{v}}{dt} = \frac{d}{dt}(\dot{r}\hat{r} + r\dot\theta\hat\theta)$$

$$= \ddot{r}\hat{r} + \dot{r}\frac{d\hat{r}}{dt} + \dot{r}\dot\theta\hat\theta + r\ddot\theta\hat\theta + r\dot\theta\frac{d\hat\theta}{dt}$$

$$= \ddot{r}\hat{r} + \underbrace{\dot{r}\dot\theta\hat\theta}_{\text{from above}} + \dot{r}\dot\theta\hat\theta + r\ddot\theta\hat\theta + r\dot\theta(-\dot\theta\hat{r})$$

$$= (\ddot{r} - r\dot\theta^2)\hat{r} + (2\dot{r}\dot\theta + r\ddot\theta)\hat\theta$$

$$= a_r\hat{r} + a_\theta\hat\theta$$

$$\frac{d\hat\theta}{dt} = \frac{d}{dt}(-\hat{i}\sin\theta + \hat{j}\cos\theta)$$

$$= -\cos\theta\,\dot\theta\,\hat{i} - \sin\theta\,\dot\theta\,\hat{j}$$

$$= -\dot\theta(\cos\theta\,\hat{i} + \sin\theta\,\hat{j})$$

$$= -\dot\theta\hat{r}$$

where we have defined $a_r := \ddot{r} - r\dot\theta^2$ as the component of the acceleration in the $\hat{r}$-direction, and $a_\theta : = 2\dot{r}\dot\theta + r\ddot\theta$ as the component of the acceleration in the $\hat\theta$-direction.

Thus, the angular momentum

$$\vec{L} = \vec{r} \times m\vec{v} = r\hat{r} \times m(\dot{r}\hat{r} + r\dot{\theta}\hat{\theta})$$

$$= mr\dot{r}\cancelto{0}{\hat{r}\times}\hat{r} + mr^2\dot{\theta}\underbrace{\hat{r}\times\hat{\theta}}_{\hat{k}}$$

$$= mr^2\dot{\theta}\hat{k}$$

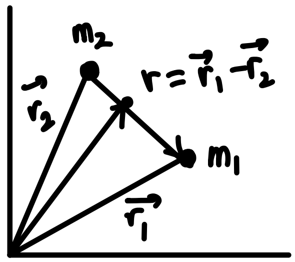
which implies that $L_z = mr^2\dot{\theta}$.

Going back to the expression for the rate at which the area is swept we hat

$$\frac{dA}{dt} = \frac{1}{2}r^2\dot{\theta} = \frac{L_z}{2m}$$ constant for any central force

$$\Rightarrow \frac{dA}{dt} = \text{constant.}$$

Central force motion as a one-body problem



An isolated system of two particles interacting under a central force $f(r)\hat{r}$

The equations of motion are: $m_1\ddot{\vec{r}}_1 = f(r)\hat{r}$ ①

$m_2\ddot{\vec{r}}_2 = -f(r)\hat{r}$ ②

$f(r)<0$: attractive
$f(r)>0$: repulsive

Let's write ① and ② in terms of $\vec{r} = \vec{r}_1 - \vec{r}_2$ and the center of mass:

$$\vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1+m_2}$$

Now $\vec{r}$: divide ① by m_1 and ② by m_2 to get

$$\ddot{\vec{r}}_1 - \dot{\vec{r}}_2 = \frac{f(r)\hat{r}}{m_1} + \frac{f(r)\hat{r}}{m_2}$$

$$\ddot{\vec{r}} = \left(\frac{m_2+m_1}{m_1m_2}\right) f(r)\hat{r}$$

Thus $\boxed{\frac{m_1m_2}{m_1+m_2}}\ddot{\vec{r}} = f(r)\hat{r} \Rightarrow \boxed{\mu\ddot{\vec{r}} = f(r)\hat{r}}$

Let's call
this the reduced mass
and denote it by μ

Now consider $\vec{R}$. add ① and ② and divide by m_1+m_2:

$$m_1\ddot{\vec{r}}_1 + m_2\ddot{\vec{r}}_2 = f(r)\hat{r} - f(r)\hat{r} = 0$$

$$\Rightarrow \frac{m_1\ddot{\vec{r}}_1 + m_2\ddot{\vec{r}}_2}{m_1+m_2} = 0$$

$$\Rightarrow \boxed{\ddot{\vec{R}} = 0}$$

So, we can now integrate this twice to obtain an equation for $\vec{R}(t)$.

$$\dot{\vec{R}}(t) = \vec{v}$$

$$\boxed{\vec{R}(t) = \vec{v}t + \vec{R}_0}$$

origin at the center of mass? $\vec{R}_0 = \vec{0}$

center of mass is stationary? $\vec{v} = \vec{0}$

★ This is an equation of motion for a single particle (It's not generalizable to systems with more than two particles).

$\vec{r}$ and $\vec{R}$ are known. Since $\vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1+m_2}$

Rearranging ⇒ $\vec{r}_1 = \left((m_1+m_2)\vec{R} - m_2\vec{r}_2\right)\frac{1}{m_1}$

and we also have $\vec{r} = \vec{r}_1 - \vec{r}_2$. Thus $\vec{r}_2 = \vec{r}_1 - \vec{r}$ which can give us

$$\vec{r}_1 = \frac{1}{m_1}\left((m_1+m_2)\vec{R} - m_2(\vec{r}_1 - \vec{r})\right)$$

$$\Rightarrow \left(1+\frac{m_2}{m_1}\right)\vec{r}_1 = \frac{m_1+m_2}{m_1}\vec{R} + m_2\vec{r}$$

$$\Rightarrow \left(\frac{m_1+m_2}{m_1}\right)\vec{r}_1 = \frac{m_1+m_2}{m_1}\vec{R} + m_2\vec{r}$$

$$\Rightarrow \boxed{\vec{r}_1 = \vec{R} + \frac{m_1 m_2}{m_1+m_2}\vec{r}}$$

and similarly $\boxed{\vec{r}_2 = \vec{R} - \frac{m_1 m_2}{m_1+m_2}\vec{r}}$ check this as an exercise

Conservation of mass (written in polar coordinates)

reduced mass again

The kinetic energy of μ is $K = \frac{\mu v^2}{2}$

Recall that we've shown that the velocity in polar coordinates is $\boxed{\vec{v} = \dot{r}\hat{r} + r\dot{\theta}\hat{\theta}}$

(see pg 56)

Thus $K = \frac{\mu}{2}(\dot{r}\hat{r} + r\dot{\theta}\hat{\theta})^2 = \frac{\mu}{2}(\dot{r}^2 + \cancel{2\dot{r}r\dot{\theta}\hat{r}\cdot\hat{\theta}} + r^2\dot{\theta}^2)$

0, since $\hat{r}\,\hat{\theta}$ perpendicular

$$= \frac{\mu}{2}(\dot{r}^2 + r^2\dot{\theta}^2)$$

There is also potential energy associated with the central force $f(r)$. For computing this let's make a few remarks first.

If the central force is a conservative force*, then the magnitude $f(r)$ of a central force can always be expressed as the derivative of a time-independent potential energy function $U(r)$

$$f(r) = -\frac{dU}{dr} \quad \Rightarrow \quad U(r) = -\int_\infty^r f(\tilde{r})\,d\tilde{r} \qquad (U \to 0 \text{ as } r \to \infty)$$

potential energy

$$\left[\quad W = \int_{\vec{r}_1}^{\vec{r}_2} \vec{f}(r)\cdot d\vec{r} = \int_{\vec{r}_1}^{\vec{r}_2} f(r)\,\hat{r}\cdot d\vec{r} = \int_{r_1}^{r_2} f(r)\,dr = \int_{r_1}^{r_2} -\frac{dU}{dr}\,dr \right.$$

work done

$$\left. = U(r_1) - U(r_2) \quad \right]$$

Thus, the total energy is given by

$$E = K + U \qquad (= \text{kinetic energy} + \text{potential energy})$$

$$= \frac{1}{2}\mu\dot{r}^2 + \frac{1}{2}\mu r^2\dot{\theta}^2 + U(r)$$

true potential

$$= \frac{1}{2}\mu\dot{r}^2 + \underbrace{\boxed{\frac{1}{2}\frac{L^2}{\mu r^2}} + U(r)}_{U_{\text{eff}}(r)}$$

centrifugal potential

here we used that the angular momentum is $\boxed{L = \mu r^2\dot{\theta}}$ (see page 56)

Thus, overall, we have
$$\boxed{E = K + U_{\text{eff}} = \frac{1}{2}\mu\dot{r}^2 + U_{\text{eff}}(r)}$$

Energy equation for a particle moving in one dimension

NOTE: all reference to θ is gone!

* In physics, a conservative force is a force with the property that the total work done in moving a particle between two points is indep. of the path taken.

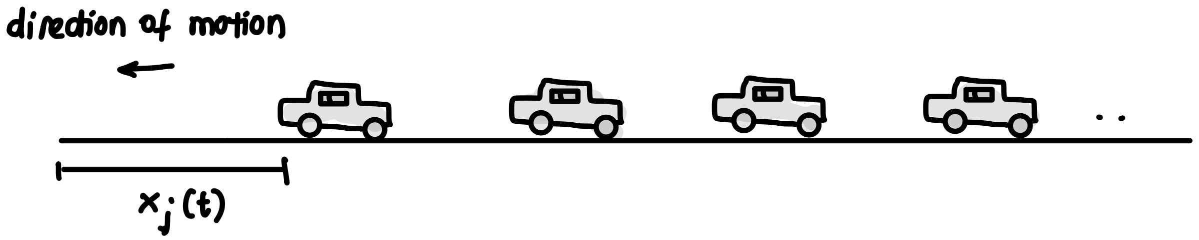
Modeling of traffic flow

Two different ways. (A) A microscopic approach based on the dynamics of single cars
(B) A mean field approach that employs an analysis on the level of fluxes and densities of vehicles.

From individual vehicles to vehicle densities.

Suppose there are N vehicles in one traffic lane, all of equal length ℓ and mass m

They are labeled $j=1,\dots,N$
↑
leading vehicle



Assumption: Vehicles cannot overtake each other

A delay differential equation for the vehicle positions

Suppose that the average values of $|x_{j+1}(t)-x_j(t)|$ are relatively small for all $j=1,\dots,N-1$
↰ distances betⁿ vehicles

→ Avoid collisions by braking when they come too close.

→ The braking force of vehicle $j+1$ will be higher, the smaller the distance $|x_{j+1}(t)-x_j(t)|$ to the jth vehicle and the faster it approaches the jth vehicle

i.e. the larger the relative velocity $\frac{d}{dt}(x_{j+1}(t)-x_j(t))$

★ The response of the driver of vehicle $j+1$ is delayed by $\tau>0$, where for simplicity we assume that the $\boxed{\text{reaction time } \tau}$ is constant for all drivers.

Braking force

$$F_{j+1}(t+\tau) = \underset{\substack{\nearrow\\ k>0\\ \text{constant}}}{k}\ \frac{\dot{x}_{j+1}(t)-\dot{x}_j(t)}{|x_{j+1}(t)-x_j(t)|}$$

Using Newton's second law of motion:

$$m \frac{d^2 x_{j+1}}{dt^2}(t+\tau) = k \frac{\dot{x}_{j+1}(t) - \dot{x}_j(t)}{|x_{j+1}(t) - x_j(t)|} = k \frac{d}{dt} \ln|x_{j+1}(t) - x_j(t)|$$

which can be integrated to yield (after we divide by m):

$$⊛ \quad \boxed{\frac{dx_{j+1}}{dt}(t+\tau) = \frac{k}{m} \ln|x_{j+1}(t) - x_j(t)| + \underset{\text{constant of integration}}{a_{j+1}}} \quad \text{for } j=1,\dots,N-1$$

system of N−1 delay differential equations (DDE)

where the position $x_1(t)$ and velocity of the first vehicle is given.

We cannot solve ⊛ analytically but we can find a numerical solution.

Densities and fluxes

The velocity of cars decreases when their density increases.

Consider a street section of length $\boxed{2s \gg L}$ and define the density of vehicles at x at time t to be

$$\rho(x,t) = \frac{\#\text{ vehicles in } (x-s, x+s) \text{ at time } t}{2s}$$

where we assume that the street section is symmetric around the position $x \in \mathbb{R}$.

We regard the density ρ as a macroscopic variable that replaces the microscopic description in terms of the positions of single vehicles by a <u>coarse-grained</u> description in terms of (average) numbers of cars per street section

We want to analyse the maximum capacity of the traffic lane under equilibrium conditions. We assume that the observed speed v of vehicles at (x,t) depends only on the density ρ. We write

$$v(x,t) = v(\rho(x,t))$$

There exist ρ_{crit} = critical density below which the vehicles move at the maximum possible speed v_{max}

ρ_{max} = maximum density at which the flow stops

From the critical to the maximum density, v decays towards zero

$$v'(\rho) \leq 0 \quad \leftarrow \text{decreasing fcn of density.}$$

Steady state and equilibrium flow

We suppose that all vehicles are separated by a distance $d>0$ and move at the same constant speed v. The equilibrium density corresponding to this situation is

$$\rho(x,t) = (d+\ell)^{-1} \qquad (x,t) \in \mathbb{R} \times [0,\infty)$$

Recall from before that $\frac{dx_{j+1}}{dt}(t+\tau) = \frac{k}{m} \ln|x_{j+1}(t) - x_j(t)| + a_{j+1}$ (DDE)

and since all vehicles move at the same speed $v_j = \frac{dx_j}{dt}$, it follows that

$$v_j = \frac{k}{m} \ln|\underbrace{x_{j+1}(t) - x_j(t)}_{\frac{1}{\rho} = \frac{1}{(1/(d+\ell))} = d+\ell}| + a_j$$

$$v_j = v, \; a_j = a \quad \Rightarrow \quad v = \frac{k}{m} \ln(d+\ell) + a$$

Notation: $\lambda = \frac{k}{m}$, $\rho = \frac{1}{d+\ell}$

$$\Rightarrow \quad v = \lambda \ln\left(\frac{1}{\rho}\right) + a$$

$$\Rightarrow \quad \boxed{V = -\lambda \ln(\rho) + a}$$

parameters to be determined from the data

From the definition of ρ_{max} it follows that $v(\rho_{max}) = 0$ which gives

$$0 = -\lambda \ln(\rho_{max}) + a$$

$$\boxed{a = \lambda \ln(\rho_{max})}$$

Thus, substituting this into $v = -\lambda \ln(\rho) + a$ we obtain

$$v = -\lambda \ln(\rho) + \lambda \ln(\rho_{max})$$

$$\boxed{v = -\lambda \ln\left(\frac{\rho}{\rho_{max}}\right)}$$

An expression for λ is easily obtained by requiring that v is continuous as a functional of ρ. Setting $\boxed{v_{max} = v(\rho_{crit})}$, we get

$$\begin{matrix} \rho = \rho_{crit} \\ v = v_{max} \end{matrix} \Rightarrow v_{max} = -\lambda \ln\left(\frac{\rho_{crit}}{\rho_{max}}\right)$$

Which gives

$$\boxed{\lambda = \frac{-v_{max}}{\ln\left(\frac{\rho_{crit}}{\rho_{max}}\right)} = \frac{v_{max}}{\ln\left(\frac{\rho_{max}}{\rho_{crit}}\right)}}$$

Altogether, we have the general relation:

$$(☺) \quad v(\rho) = \begin{cases} v_{max}, & \rho \le \rho_{crit} \\ -\frac{v_{max}}{\log\left(\frac{\rho_{max}}{\rho_{crit}}\right)} \ln\left(\frac{\rho}{\rho_{max}}\right) = \frac{v_{max}}{\ln\left(\frac{\rho_{max}}{\rho_{crit}}\right)} \ln\left(\frac{\rho_{max}}{\rho}\right), & \rho > \rho_{crit} \end{cases}$$

Maximum traffic flux at equilibrium. We define the instantaneous traffic flux J as the # of vehicles passing through a street sector $[x, x+\Delta x)$ in the time interval $[t, t+\Delta t)$, $J = \left(\frac{\text{\# vehicles at time } t}{\Delta x}\right)\left(\frac{\Delta x}{\Delta t}\right)$

Letting $\Delta x, \Delta t \to 0$, we get

$$\boxed{J(\rho) = \rho v(\rho)}$$ density flux J

With (!!) we have

$$J(\rho) = \begin{cases} \rho v_{max} & , \quad \rho \leq \rho_{crit} \\ \dfrac{\rho v_{max}}{\log\left(\dfrac{\rho_{max}}{\rho_{crit}}\right)} \ln\left(\dfrac{\rho_{max}}{\rho}\right), & \rho > \rho_{crit} \end{cases}$$

which can be shown to attain its maximum at $\boxed{\rho^* = \frac{\rho_{max}}{e}}$

Traffic jams and propagation of perturbations

We want to study what happens when the first vehicle brakes

⤳ effect of a perturbation of the lead vehicle on the pursuing vehicles

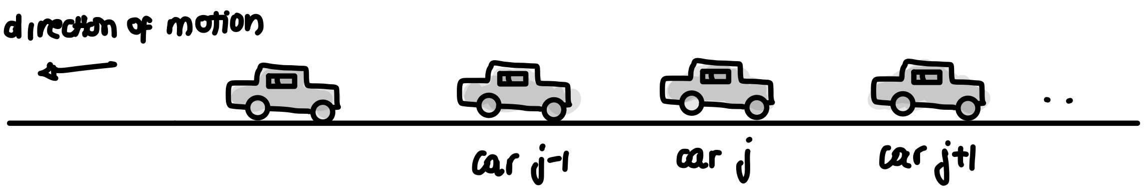
We go back to the microscopic picture again and consider a platoon of cars under maximum flux conditions. We suppose that all vehicles move at constant speed

If $\rho > \rho_{crit}$, $v(\rho) = \dfrac{v_{max}}{\ln\left(\dfrac{\rho_{max}}{\rho_{crit}}\right)} \ln\left(\dfrac{\rho_{max}}{\rho}\right)$.

If $\rho = \rho^* = \frac{\rho_{max}}{e}$, then $v(\rho^*) = \dfrac{v_{max}}{\ln\left(\dfrac{\rho_{max}}{\rho_{crit}}\right)} \cancelto{1}{\ln\left(\dfrac{\rho_{max}}{\left(\frac{\rho_{max}}{e}\right)}\right)} = \dfrac{v_{max}}{\ln\left(\dfrac{\rho_{max}}{\rho_{crit}}\right)}$

Let's assume further that we can extend the time $t \geq 0$ to the whole real axis, $t \in \mathbb{R}$, and that the lead vehicle crosses the origin $x=0$ at time $t=0$, i.e. $\boxed{x_1(0)=0}$

With the sign convention $\boxed{x_{j-1} - x_j \geq l > 0}$



and $v^* = v(\rho^*)$ we have $\dfrac{d x_{j+1}(t+\tau)}{dt} = \lambda \ln|x_{j+1}(t) - x_j(t)| + a$

with $\lambda = \frac{v_{max}}{\ln\left(\frac{\rho_{max}}{(\rho_{max}/e)}\right)} = \frac{v_{max}}{\ln(e)} = v_{max}$

$v = -\lambda\ln(\rho) + a \qquad v(\rho_{max}) = 0 \Rightarrow$

$a = \lambda\ln(\rho_{max})$

$v = -\lambda\ln(\rho) + \lambda\ln(\rho_{max}) = -\lambda\ln\left(\frac{\rho}{\rho_{max}}\right)$

$\lambda = -v/\ln(\rho/\rho_{max})$

and $a = \lambda\ln(\rho_{max}) = v_{max}\ln(\rho_{max})$

subst. $\rho = \rho_{max}$ and $v = -\lambda\ln(\rho) + a$

$v = 0$

$0 = -\lambda\ln(\rho_{max}) + a$

$a = \lambda\ln(\rho_{max})$

with $\lambda = v_{max}$

$$\Rightarrow \frac{dx_{j+1}}{dt}(t+\tau) = v_{max}\ln|x_{j+1}(t) - x_j(t)| + v_{max}\ln(\rho_{max})$$

$$= v_{max}\left[\ln(\underbrace{x_j(t) - x_{j+1}(t)}_{\text{since } x_j - x_{j+1} > 0}) + \ln(\rho_{max})\right]$$

$$= v_{max}\ln\left(\rho_{max}(x_j(t) - x_{j+1}(t))\right)$$

Breaking of the lead vehicle and perturbation of the pursuing vehicles

For $t>0$, we consider the DDE system

$$\frac{dx_1}{dt} = \phi(t) \leftarrow \text{first one behaves differently because it brakes!}$$

$$\frac{dx_j}{dt}(t+\tau) = v_{max}\ln\left(\rho_{max}(x_j(t) - x_{j+1}(t))\right) \qquad j = 2, \dots, N$$

Where we assume that the system is in equilibrium for $\boxed{t \le 0}$

$$x_j(t) = \underset{\substack{\uparrow \\ \text{model parameter}}}{v^*}t - (j-1)(d+\ell) \quad \text{for } j = 1, \dots, N$$

model parameter (not the instantaneous velocity of individual vehicles given by $v_j = \frac{dx_j}{dt}$)

We assume that the 1st vehicle with position x_1 brakes at $x=0$ and releases the break after a short time $t_b > 0$. This can be written as

$$\phi(t) = \begin{cases} v^* & t \le 0 \\ v^*(1-b(t)) & t > 0 \end{cases}$$

equilibrium speed

$$\frac{dx_1}{dt} = \phi(t) = \begin{cases} v^* & t \le 0 \\ v^*(1-b(t)) & t > 0 \end{cases}$$

↑ speed decreases from v^* according to $b(t)$

where we use $b(t) = kte^{-(t-t_b)/t_b}$

↑ decay rate

Solving the ODE for x_1, by integrating wrt time we obtain

$$x_1(t) = v^*t - v^*\int_0^t kse^{-(s-t_b)/t_b}\,ds \quad , \; t > 0$$

Integrating by parts we get $u = s \quad \frac{dv}{ds} = e^{-(s-t_b)/t_b}$

$\frac{du}{ds} = 1 \quad v = -t_b e^{-(s-t_b)/t_b}$

$$x_1(t) = v^*t - v^*k\left[-s\,t_b e^{-(s-t_b)/t_b}\right]_0^t - v^*k\int_0^t t_b e^{-(s-t_b)/t_b}\,ds$$

$$= v^*t - v^*k\left(-t t_b e^{-t/t_b} e\right) + v^*k\,t_b^2\left[e^{-(s-t_b)/t_b}\right]_0^t$$

$$= v^*t + v^*k t t_b e^{-\frac{t}{t_b}} e + v^*k t_b^2 e^{-t/t_b} e - v^*k t_b^2 e$$

$$= v^*t + ev^*k t_b\left[t e^{-t/t_b} + t_b e^{-t/t_b} - t_b\right]$$

$$= v^*t + ev^*k t_b\left[(t+t_b)e^{-t/t_b} - t_b\right]$$

We call $y_j(t)$ the hypothetical position of the jth car, if the lead vehicle had not braked, i.e. without the perturbation

We also define the perturbation displacement due to the perturbation of the lead vehicle's motion:

$$z_j(t) = \underbrace{x_j(t)}_{\text{true position}} - \underbrace{y_j(t)}_{\text{hypothetical position}}$$

The perturbation displacement of the first vehicle then is

$$z_1(t) = \cancel{v^*t} + ev^*k t_b\left[(t+t_b)e^{-t/t_b} - t_b\right] - \cancel{v^*t}$$

y_j = equilibrium position

which is → $= -v^*\int_0^t b(s)ds, \; t>0$

By $x_j(t) = v^*t - (j-1)(d+l)$, $j=1,\dots,N$, it follows that the pursuing vehicles with $j=2,\dots,N$ satisfy

$$z_j(t) = x_j(t) - \left(v^*t - (j-1)(d+l)\right) = x_j(t) - v^*t + (j-1)(d+l), \; t>0$$

Note that $\boxed{z_j(t)=0 \text{ for } t\leq 0}$ and for all $j=1,\dots,N$. Further note that the non-collision constraint

$$x_{j-1}(t) - x_j(t) > l \quad \forall t \in \mathbb{R}$$

at least the length of the car

implies that

$$z_j(t) - z_{j-1}(t) = x_j(t) - \cancel{v^*t} + (\cancel{j}-1)(d+l) - x_{j-1}(t) + \cancel{v^*t} - (\cancel{j}-2)(d+l)$$

$$= x_j(t) - x_{j-1}(t) + d + l$$

Upon rearrangement

$$\Rightarrow l < x_{j-1}(t) - x_j(t) = z_{j-1}(t) - z_j(t) + d + l$$

$$z_{j-1}(t) - z_j(t) > \cancel{l} - \cancel{l} - d$$

$$\boxed{z_j(t) - z_{j-1}(t) < d} \quad \forall t \in \mathbb{R}.$$

Reaction time and the onset of traffic jam

These new equations allow us to recast the DDE (delay-differential equations) system

$$\frac{dx_1}{dt} = \phi(t)$$

$$\frac{dx_j}{dt}(t+\tau) = v^* \ln\left(\rho_{max}(x_{j-1}(t) - x_j(t))\right) \quad j=2,\dots,N \qquad (1)$$

as a DDE for the perturbation displacement z_j

Recall that we showed $\frac{\rho_{max}}{e} = \rho^* = \frac{1}{d+l} \Rightarrow \boxed{d+l = \frac{e}{\rho_{max}}}$ under the maximum flow conditions. This implies that the pursuing vehicles have a perturbation displacement that satisfies

$$z_j(t) = x_j(t) - \left(v^*t - (j-1)(d+l)\right)$$

$$= x_j(t) - v^*t + (j-1)\frac{e}{\rho_{max}}, \quad t > 0 \qquad (2)$$

Differentiating (2) wrt t: $\frac{dz_j}{dt} = \frac{dx_j}{dt} - v^*$

$$\frac{dx_j}{dt} = \frac{dz_j}{dt} + v^*$$

If we evaluate this at $t = t + \tau$ and subst. this & (2) into (1) then we obtain

$$\frac{dz_j}{dt}(t+\tau) = -v^* + v^* \ln\Big[\rho_{max}\Big(z_{j-1}(t) + \cancel{v^*} t - (\cancel{j}-2)\frac{e}{\rho_{max}} - z_j(t) - \cancel{v^*} t + (\cancel{j}-1)\frac{e}{\rho_{max}}\Big)\Big)$$

$$= v^* \ln\left(\rho_{max}\left(z_{j-1}(t) - z_j(t) + \frac{e}{\rho_{max}}\right)\right) - v^*$$

for $j=2,\dots,N$. With the lead vehicle displacement $z_1(t) = -v^*\int_0^t b(s)\,ds$

and initial conditions $z_j(0)=0$, $j=2,\dots,N$.

Lecture 17

Some problems in probability

(Notes by Percy Deift, NYU)

Probabilistic reasoning is often very different from the kind of reasoning we meet and employ in everyday life. Increasingly we are presented in the news, in newspapers, in the internet and on television with statistical figures and tables. But statistics is based on probability theory and so it is important for us to understand basic probability theory.

Some notation:

① A set is a collection of objects which we usually denote by a capital letter e.g X or Y. We will mostly consider finite sets, so $X = \{x_1, x_2, \dots, x_n\}$, $n<\infty$ where the x_i's are elements with the following 2 properties.

(a) $P(X)=1$

and (b) $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ ← empty set

Note that it follows from (b) that if $\{x_i\}$ is the singleton set containing only the element x_i, $P_i \equiv P(\{x_i\})$ then

(c) $P(A) = \sum\limits_{x_i \in A} P_i$

We think of all sets $A \subset X$ as events: thus $P(A)$ is the probability that event A happens.

(a) means that the full event X is meant to happen

(b) If $f: X \to \mathbb{R}$ is a function from X to $\mathbb{R}$ then the average of f, or the expectation of f is given by $\mathbb{E}(f) = \sum\limits_{i=1}^{n} f(x_i)\, p_i$

⌈ From wiki: Consider a random variable X with a finite list $x_1, x_2, \dots, x_k$ of possible outcomes, each of which has probability $p_1, \dots, p_k$ of occuring. Then the expectation of X is defined as

$$\mathbb{E}(X) = x_1 p_1 + x_2 p_2 + \dots + x_k p_k$$

since $\sum p_i = 1$ it is natural to interpret $\mathbb{E}(X)$ as a weighted average of the x_i values with weights given by their probabilities p_i.

(i) Coin tossing

Here X has two elements $x_1 = H$, $x_2 = T$. We say that the coin is fair if $P_H = P(\{H\}) = \frac{1}{2}$ and $P_T = P(\{T\}) = \frac{1}{2}$

Suppose one wins a dollar if he throws a heads and nothing if one throws a tails. Then let $f(H) = 1$, $f(T) = 0$ & $\mathbb{E}(f) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}$

(ii) Throwing a die

Here X has 6 elements, $x_1 = 1, x_2 = 2, \dots, x_6 = 6$

Again the die is fair if $p_i = P(\{x_i\}) = \frac{1}{6}$

If $A = \{2, 4, 6\}$ is the event that we obtain an even number then

$$P(A) = P_2 + P_4 + P_6 = \frac{1}{6}(3) = \frac{1}{2}$$

is the probability that we obtain an even number after a throw of a die.

Our first example which demonstrates that probabilistic reasoning can be very counter-intuitive is the following.

Summer has arrived, school is out and a bunch of friends — there are 9 of you - want to go together to a baseball game.

Should you go to an afternoon or an evening game?

Let us assume for simplicity that on any given day, a person is free in the afternoon or the evening (but not both!) with equal probability. A text message is then sent around to all 9 friends, starting on Aug 1, say, with the following 10 questions:

On Aug 1, are you free in the afternoon or the evening?
On Aug 2, ...
⋮
On Aug 10, are you free in the afternoon or the evening?

Question : What is the probability that on one of these 10 days everybody will be free at the same time?

Guesses?

In order to analyze the problem, we note first that on any given day, when one collects the responses from the 9 friends there are $2^9 = 512$ possible outcomes.

(1) AEEAAAEEA (9 responses

(2) AAAEEAEAE

⋮

(2^9) EAAAAEEEAA

Of these outcomes only <u>two</u> are favorable:

all A's AA...A

OR all E's EE... E

Thus, the probability of success on the first evening Aug 1 is

$$\frac{2}{2^9} = \frac{1}{2^8} = \frac{1}{256}$$

Now, the key to analyzing the problem is to consider the probability of failure rather than success. If A is the event "success" and B is the event "failure", then clearly $A \cap B = \phi$ and so $P(A \cup B) = P(A) + P(B) = \frac{1}{256} + P(B)$.

But $A \cup B = X$, the full set and so $P(B) = 1 - \frac{1}{256} = \frac{255}{256}$

Now, what happens on Aug 2 is <u>independent</u> of Aug 1 and so the probability of failure on Aug 1 and Aug 2 is just

$$P(B)P(B) = \left(\frac{255}{256}\right)^2$$

and continuing we see that after 10 days the probability of failure on all 10 days is given by $\left(\frac{255}{256}\right)^{10} \approx 0.9616$

Thus the probability of success after 10 days is less than 0.04 is 4%!

In order to have more than a 50% success you'd have to offer 178 consecutive options from Aug 1 till some time in February when the season is over.

$$\ulcorner \left(\frac{255}{256}\right)^n = 0.5 \rightarrow \log_{\frac{255}{256}}(0.5) = n \doteq 178 \lrcorner$$

If you offered 365 days, 1 year of options, your chance of success is about 75%.

So if you want to go to a game, or a movie, with a large group of friends, just fix a day and stick with it!

Example 2

Let us consider the birthday problem

Question. If I offered you a bet that two people in this class have the same birthday would you take the bet?

To win you would certainly want at least a 50% chance of winning. We can work out the odds in the following way.

If there is only one person in the class there is clearly no problem. So suppose there are two people in the class. Again the trick is to consider the probability that they do not have the same birthday. Then the first person has his or her birthday on any one of 365 days. But then the other person must have his/her birthday on one of the remaining 364 days. There are 365 x 365 ways for the 2 birthdays to occur, so the probability they do not have a common birthday is

$$\frac{365 \times 364}{365 \times 365}$$

Hence the probability that they have a common birthday is

$$1 - \frac{365 \times 364}{365 \times 365} = 1 - \frac{364}{365} = \frac{1}{365} = 0.002 = 0.2\%$$

Now suppose there are 3 people in the class. Then the probability that they have a common birthday is $1 - \frac{365 \times 364 \times 363}{(365)^3} = 1 - 0.991 = 0.09 = 0.9\%$

More generally if there are n people in the class then the probability that 2 have the same birthday is

$$q_n = 1 - \frac{365 \times 364 \times \dots \times (365-n+1)}{(365)^n}$$

⌈e.g. for $n=3$ we have $1 - \frac{365 \times 364 \times 363}{(365)^3}$ ⌋

We can write q_n more compactly as

$$q_n = 1 - \frac{365\,!}{(365-n)!\; 365^n}$$

where $x! = x(x-1)(x-2)\cdots 1$

We find for $n=10$ $q_{10} \sim 0.117 \approx 11\%$

$q_{20} \sim 0.412 \simeq 41\%$

$q_{30} \sim 0.707 \simeq 70\%$

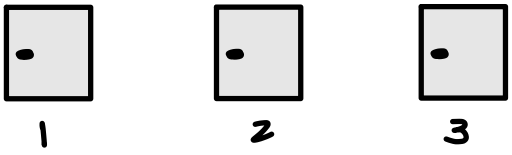
So where is the break point when you have a 50% chance of winning?

If $\boxed{n=23}$ then $q_{23} = 0.508 \simeq 50\%$

Let's see how this works out in our class...

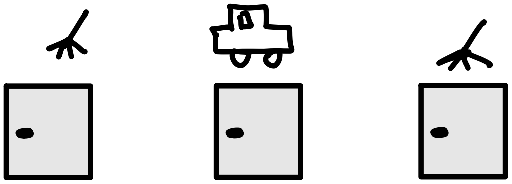
Example 3

This problem was made famous on the Monty Hall television show, "Let's make a deal!" The game works in the following way, the host Monty shows a player 3 doors on the stage



Hidden behind one of the doors is a valuable prize, e.g a car but hidden behind the other two doors are "gags" e.g. broomsticks.

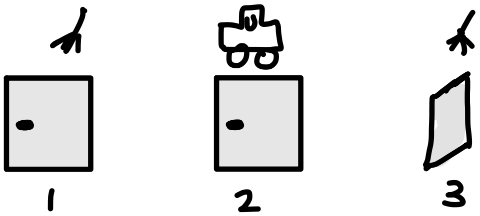
e.g.



The player chooses but does not open the door Monty who knows where the car is then opens one of the doors concealing a broom for the player to see.

For example, if the player chose door 2, Monty would open either door 1 or door 3. As there are 2 brooms there will always be at least one door with a broom behind it.

So suppose he opens door 3

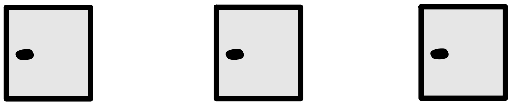


Monty then asks the player if he wants to switch from door 2 to door 1 in this case.

Question. Should he/she switch? What do you think?

Most people think it doesn't help to switch, the odds are 50/50. But it turns out on a more careful analysis that there is a distinct advantage to switch.

To see how this works, consider the following. For the 3 doors, there are at the outset, precisely 3 possible configurations of the brooms & the car



	1	2	3
(i)	C	B	~~B~~
(ii)	B	C	~~B~~
(iii)	B	~~B~~	C

Now suppose the player chooses door 1. The same argument works for 2 or 3
Then for config. (i) Monty opens door 2 or door 3, say door 3.

For config. (ii) Monty opens door 3 & for config. (iii) he opens door 2.
Now the situation is clear: the player is being offered to change his choice to a door with the following property: for 2 of the configurations (i), (ii), or (iii) the remaining door contains a car and only for one there is a broom.
Thus, he has a $\frac{2}{3}$ chance of winning the car if he switches, but only a $\frac{1}{3}$ chance if he does <u>not</u> switch.

Everyone would agree that this situation is counterintuitive to everyday reasoning but the probabilistic reasoning is irrefutable.

Lecture 18

Example 4

All the problems considered so far have involved finding the right approach but the mathematics involved was rather simple.
In the next problem, the math will be more substantial:

The problem Suppose that in a certain month bad things happen to you at least 3 days in a row. Is someone out to get you, or is it just in the cards?

To analyze this problem we make the following simplifying assumptions:

→ with probability $\frac{1}{2}$ a day is good & with probability $\frac{1}{2}$ a day is bad

<u>Specific question</u>: What is the probability that in a given month, you have (at least) 3 bad days in a row?

Guesses?

NOTATION A bad month is a month in which we have (at least) 3 bad days in a row. So what is $P(\{\text{bad month}\})$?

More **NOTATION**: We denote a bad day with a 1
a good day with a 0

To get some feeling for the problem, consider a sequence of 5 consecutive days — a 5-sequence. We say a 5-sequence, or more generally an n-sequence is bad if it contains (at least) 3 bad days in a row. otherwise we say the n-sequence is good

Now there are clearly $2^5 = 32$ different 5-sequences (see next page for what a, b, c, d denote)

bad	11111	bad	10111	bad	01111	bad	00111
bad	11110	d	10110	bad	01110	d	00110
bad	11101	b	10101	b	01101	b	00101
bad	11100	c	10100	c	01100	c	00100

same endings within each of these 4 groups

a	11011	a	10011	a	01011	a	00011
d	11010	d	10010	d	01010	d	00010
b	11001	b	10001	b	01001	b	00001
c	11000	c	10000	c	01000	c	00000

same endings within each of these 4 groups

Thus $8/32$ sequences are bad 5-sequences: so $P(\{\text{bad 5-sequences}\}) = \frac{8}{32} = \frac{1}{4}$

Want to compute $P(\{\text{bad } n\text{-sequence}\})$ for any n, in particular for $n = 30$ days $= 1$ month.

How do we proceed?

Notice that every n-sequence either ends with

$\cdots\ 11$
$\cdots\ 01$
$\cdots\ 00$
$\cdots\ 10$

$2^2 = 4$ combinations

Let

$a_n \equiv \#\{\text{good } n\text{-sequences ending in } 11\}$
$b_n \equiv \#\{\text{good } n\text{-sequences ending in } 01\}$
$c_n \equiv \#\{\text{good } n\text{-sequences ending in } 00\}$
$d_n \equiv \#\{\text{good } n\text{-sequences ending in } 10\}$

count # of a, b, c, d sequences in previous page

For n=5 we see $a_n = 4$
$b_n = 7$
$c_n = 7$
$d_n = 6$

→ Note $4+7+7+6 = 24$ good 5-sequences
$+\ 8$ bad 5-sequences
32 ✓

Now comes the crucial step. Consider a_{n+1}, the # of good $(n+1)$-sequences ending in 11. Such a sequence must look like

$\cdots 011$

but not $\cdots 111$ ← bad sequence:
or $\cdots 001$
or $\cdots 101$

Thus $\boxed{a_{n+1} = b_n}$
$\uparrow\ \cdots 11 \qquad \uparrow\ \cdots 01$

① $b_n = \#\{\text{good 5-sequence ending in } 01\}$ e.g 11001
$a_{n+1} = \#\{\text{good 6-sequence ending in } 11\}$ e.g. $\underbrace{11001}_{b_n}1$

Now consider $b_{n+1} = \#\{\text{good } n{+}1\text{-sequences ending in } 01\}$

Such a sequence must look like

↗ e.g. $10\,\underline{0001}$

$c_n = \#\{\text{good 5-sequence ending in } 00\}$

$d_n = \#\{\text{good 5-sequence ending in } 10\}$

↳ e.g. $110\underline{101}$

$\dots 101$

or $\dots 001$

Thus $\boxed{b_{n+1} = \underset{\dots 00}{c_n} + \underset{\dots 10}{d_n}}$ ②

Similarly for c_{n+1}: $\dots 100$

$\dots 000$

$\Rightarrow \boxed{c_{n+1} = c_n + d_n}$ ③

and for d_{n+1}: $\dots 010$

$\dots 110$

$\Rightarrow \boxed{d_{n+1} = b_n + a_n}$ ④

We can write ①, ②, ③, ④ in matrix form

$$\begin{pmatrix} a_{n+1} \\ b_{n+1} \\ c_{n+1} \\ d_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_n \\ b_n \\ c_n \\ d_n \end{pmatrix} \qquad ⑤$$

or if we let

$$X = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad x_n = \begin{pmatrix} a_n \\ b_n \\ c_n \\ d_n \end{pmatrix} \Rightarrow \boxed{x_{n+1} = X x_n} \quad n \geqslant 2$$

Iterating $x_n = X x_{n-1} = X(X x_{n-2})$

$$= X^2 x_{n-2}$$

$$= X^3 x_{n-3} \quad \ldots$$

$$= X^{n-2} x_2 \qquad ⑥$$

Clearly $x_2 = \begin{pmatrix} a_2 \\ b_2 \\ c_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

a_2 = # {good 2-sequences ending in 11}

11 one of each
10
01
00

with $|x_n| = a_n + b_n + c_n + d_n$ = # {good n-sequences}

Then $P(\{\text{good } n\text{-sequences}\}) = \frac{|x_n|}{2^n}$ ← total number of combinations. if $n=5$, $2^n = 32$.

$P(\{\text{bad } n\text{-sequences}\}) = 1 - \frac{|x_n|}{2^n}$

Now we can simplify the system by noting from ② and ③ that

$$b_{n+1} = c_n + d_n = c_{n+1} \qquad n \geq 2$$

But $b_2 = c_2 = 1$ and so $b_n = c_n$ for all $n \geq 2$

Thus, our equations ① – ④ take the form

$$\left[\begin{array}{l} a_{n+1} = b_n \\ b_{n+1} = b_n + d_n \\ d_{n+1} = b_n + a_n \end{array} \right]$$

or in matrix form $y_{n+1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} y_n \quad n \geq 2$ where $y_n = \begin{pmatrix} a_n \\ b_n \\ d_n \end{pmatrix}$

Again, $y_2 = \begin{pmatrix} a_2 \\ b_2 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

i.e. $y_{n+1} = Y y_n$ for $Y = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ and as before $y_n = Y^{n-2} y_2$ ⑦

Check. Take $n=5$ Now $Y^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$

$$Y^3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \\ 1 & 3 & 2 \end{pmatrix}$$

Thus $y_5 = Y^3 y_2 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 4 & 2 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 6 \end{pmatrix}$

$a_5 = 4$

$c_5 = b_5 = 7$ $\qquad |x_5| = 4 + 2 \times 7 + 6 = 24$

$d_5 = 6$

as before! ☺

and hence $P(\{\text{good 5-sequence}\}) = \frac{4+14+6}{32} = 0.75$

$P(\{\text{bad 5-sequence}\}) = \frac{8}{32} = 0.25$

We are interested in $y_{30} = Y^{28}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = (Y^7)^4 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and putting this on a computer or if you have the power just doing by hand we find

$$|x_{30}| = a_{30} + 2b_{30} + d_{30}$$

$$P(\{\text{good month}\}) = \frac{|x_{30}|}{2^{30}}$$

$$P(\{\text{bad month}\}) = 1 - \frac{|x_{30}|}{2^{30}} = 0.907$$

Thus the probability of at least 3 bad days in a row in a month is over 90%, which is pretty high. So don't think anyone is out to get you if too many bad things go wrong in a row. It's just the way it is.

The good, but perhaps counterintuitive, news is that

$$P(\{\text{at least 3 good days in a row}\})$$

is also 0.907 ~ 90%. So in any given month we can expect some good stretches. But somehow, our psychology is such that we don't remember them as vividly as the bad stretches.

Lecture 19

The mathematics of voting, power, and sharing

(Ian Griffiths notes)
University of Oxford

VOTING SYSTEMS

A voting system is a way for a group of people to select one from among several possibilities.

If only 2 alternatives then it's easy → alternative that is preferred by the majority wins.
(difficulty arises if there is a tie)

When several people have to choose among more than two alternatives than things get trickier

Simple example showing one of the oldest voting paradoxes.

Suppose a group of say 60 people will meet for a celebration in a restaurant, and the restaurant manager wants them to pick one menu for the whole group.

Main course choices : salmon
or chicken

The organizers consult their group & find that the majority prefers salmon.

The owner later calls up & says that her fish supplier has become less reliable & she is now offering a choice between chicken & beef.

The group now is consulted again and prefers the chicken choice.

In summary, the group
- prefers salmon over chicken
- prefers chicken over beef.

A day later, the restaurant managers calls back; she has switched to another supplier and she can again offer salmon

However, the Department of Agriculture recently destroyed large quantities of chicken because of a microbial contamination and the choice is now between salmon and beef.

The organizers feel sure, in view of the ranking above, that their group will largely prefer salmon, but when they ask, they find a clear majority for beef.

The group prefers beef over salmon.

"Oh well," they think. "people are fickle, some of them must have changed their minds". Yet, this was not the case: every single person polled had a clear ranking of the 3 possibilities and stuck to that ranking in a consistent way. Nonetheless, even though every single individual is entirely consistent, the group is not.

We'll now look at a numerical example.

Suppose that

∘ 25 people rank	20 people rank	15 people rank
1. salmon	1. chicken	1. beef
2. chicken	2. beef	2. salmon
3. beef	3. salmon	3. chicken

So,	salmon > chicken 25 + 15 = 40	chicken > beef 25+20 = 45	beef > salmon 20 + 15 = 35

The paradoxical behavior of the group is explained.

This kind of paradox happens all the time and for things more serious than this, such as presidential elections.

In the case where this type of paradox doesn't happen, that is, when there is one alternative that is always preferred by a majority (although not always the same majority) if it were in a one-on-one race against any one of the others, then we call the winning alternative the "Condorcet" winner [this would be the case for the "chicken" choice in the example above if the third group had changed their ordering to

1. beef
2. chicken
3. salmon.]

E.g.

• 25 people rank	• 20 people rank	• 15 people rank
1. Salmon	1. chicken	1. beef
2. chicken	2. beef	2. chicken
3. beef	3. salmon	3. salmon.

Salmon & chicken options	chicken & beef options	salmon & beef options
salmon> chicken 25	chicken> beef 25+20=45	beef> salmon 20+15=35

* Majority prefers chicken!

We have just seen that there doesn't always exist a Condorcet winner. But when there exists one, it seems fair that that should be the winning choice for the whole group. Or does it?

Different systems to select the "winner".

Because the Condorcet method doesn't always yield a winner, it is not used a lot.

- **PLURALITY**: The candidate who is ranked in first place most often, wins. This is the way in which members of congress are elected in the U.S. in every state.

- **PLURALITY WITH RUN-OFF**: The two candidates with the most first places are retained, and then a second round run-off election is held between them. This is the system used in the election of the president of France.

- **SEQUENTIAL RUN-OFF / HARE SYSTEM**: The candidate w/ the fewest first places is removed, then (after her/his votes have been restributed among the remaining candidates) the next-bottom candidate, and so on... This system has been used for years in Australia, Ireland, and in NYC (although not in situations where only one winner has to be selected, but where several seats are available)

- **BORDA COUNT** : If there are N candidates, then every voter gives N points to his/her first, N-1 to the second choice, ...

The points that all the voters gave are then added, and the candidate with the most points wins. This system is often used in clubs to decide on admission (or not) of new members.

Different methods can lead to different outcomes.

Some paradoxical situations with a few more examples.

Example. PARADOX w/ (RUN-OFF or) SEQUENTIAL RUN-OFF

A student asks 17 of her friends what kind of breakfast they prefer. Here are the answers.

# of people for each ranking	6	5	4	2
cereal	1	2	3	2
danish	2	3	1	1
bagel	3	1	2	3

First we get rid of the alternative that got fewest first places: ~~bagel~~ (which had 5)
[danish had 4+2=6., cereal had 6]

That leaves cereal & danish.

With only these two alternatives remaining, the preferences are.

	6	5	4	2
cereal	1	1	2	2
danish	2	2	1	1

← cereal wins
because it
has the most
1st places now
(6+5 vs 4+2)
=11 =6.

But if the last group of ② votes changes its mind and decides to rank cereal above danish instead of the other way around, what happens then? Surely cereal's chances of winning must be better now?. Let's see

	6	5	4	2
cereal	1	2	3	1
danish	2	3	1	2
bagel	3	1	2	3

The item with the fewest 1st places is now the danish (4 versus 5 for bagel & 6+2=8 for cereal)

Reassigning the danish's votes we get

	6	5	4	2
cereal	1	2	2	1
bagel	2	1	1	2

people prefering cereal = 6+2=8

" bagel = 5+4 =9

So the bagel wins and cereal loses even though more voters prefered cereal than before ...

Example **PARADOX W/ BORDA COUNT**

A club of 25 people are planning an outing. They have narrowed down the choices to a trip to the beach, a hike in the mountains, or a day in San Francisco. Their preference schedule is the following

	13	10	2
beach	2	1	3
mountains	3	2	1
SF	1	3	2

This is in fact a case where there is a Condorcet winner: in the one-on-one contests SF always wins:

- beach vs SF : $13+2=15$ prefer SF
  10 prefer beach
- mountains vs SF : 13 prefer SF
  $10+2=12$ prefer mountains

SF also wins the plurality vote and is also the winner under the run-off scheme. In a Borda count, we find the following totals of points

$$\begin{aligned}\text{beach} &= (10\times3)+(13\times2)+(2\times1)\\ &= 30+26+2\\ &= 58\end{aligned}$$

WINS!

$$\begin{aligned}\text{mountains} &= (2\times3)+(10\times2)+(13\times1)\\ &= 6+20+13\\ &= 39\end{aligned}$$

$$\begin{aligned}\text{SF} &= (13\times3)+(2\times2)+(10\times1)\\ &= 39+4+10\\ &= 53\end{aligned}$$

$N=3$ candidates
1st place → 3 pts
2nd place → 2 pts
3rd place → 1 pt

This does not lead to the same winner, even though SF won by several other methods.

Lecture 20

THE POWER INDEX

In the previous lecture, all voters had equal standing. This is not true in all voting situations, as shown by the following examples.

Examples

1) SHAREHOLDERS: their vote is proportional to the number of shares they hold
2) ELECTORAL COLLEGE: many states require that their delegates vote for the same presidential candidate; as a result, states function like voters with unequal weights, and thus unequal importance in the end result.

3) COUNTY BOARDS: some townships have more representatives than others. Assuming that they all vote the same way, this gives different townships unequal power

How can one measure this power? It is not simply proportional to the number of votes:

Example: In a shareholders' meeting, there are 3 participants.

A has 47% of the shares
B has 48% ″
C has the remaining 5%.

A majority of 51% is needed to pass any measure. Any group of 2 can force the measure to pass over the opposition of the third. So A, B, C have equal power — despite their unequal number of shares

There exist several schemes to try to measure this "power" of the participants.

One of the most widely accepted is the Banzhaf power index

Motivation: When do you have "power"? When your decision matters!
That is, when whether you vote one way or the other makes a difference in the outcome or, when your vote is a "swing" vote.

So let us define your power index as the fraction

$$\boxed{\frac{\text{number of coalitions where you are a swing vote}}{\text{total number of coalitions}}}$$

Example. • In the case above (A: 47%, B: 48%, C: 5%) the possible coalitions are

1. ABC |
2. AB | C
3. AC | B
4. BC | A
5. A | BC
6. B | AC
7. C | AB
8. | ABC

In cases 1,8 : nobody is a swing vote
In cases 2,7 : A, B are both swing votes
In cases 3,6 . A, C are both swing votes
In cases 4, 5, B, C are both swing votes

It follows that A, B, and C have the same power index $\frac{4}{8} = 0.5$

- Whether you are a swing vote or not depends not only on your number of shares, but also on what majority is needed to reach a decision.

If a measure can be passed in the example above only when it has 53% of the votes or more, then the situation changes

Notation

$$[q : w_1, w_2, w_3 \ldots] = (p_1, p_2, p_3, \ldots)$$

quorum needed to pass a measure

weights of the voters

power indices of the voters

In the example above : $[51 : 47, 48, 5] = (0.5, 0.5, 0.5)$

Example $[51 : 40, 30, 20, 10] = (?, ?, ?, ?)$

4 "shareholders" : A, B, C, D

$2^4 = 16$ combinations

either + or −

circle votes that are swing votes

	A 40	B 30	C 20	D 10	Votes	Pass / Fail
1.	+	+	+	+	100	P
2.	⊕	+	+	−	90	P
3.	⊕	⊕	−	+	80	P
4.	+	−	⊕	+	70	P
5.	−	⊕	⊕	⊕	60	P
6.	⊕	⊕	−	−	70	P
7.	⊕	−	⊕	−	60	P
8.	+	−	−	+	50	F
9.	−	+	+	−	50	F
10.	−	+	−	+	40	F
11.	−	−	+	+	30	F

12.	+	–	–	–	←	40	F
13.	–	+	–	–	←	30	F
14.	–	–	+	–	←	20	F
15.	–	–	–	+	←	10	F
16.	–	–	–	–	←	0	F

$$\Rightarrow [51 : 40, 30, 20, 10] = \left(\frac{4}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$$

FAIR DIVISION

Examples . • splitting a cake
- dividing up an estate among their heirs
- splitting up the assets when a company breaks up

TWO PLAYERS (division of a "cake" between 2 people)

One cuts, the other chooses

Implicit assumptions :
- each player is able to divide cake in such a way that either of the two pieces would be OK with that player
- given any division of the cake, each player would find at least one piece acceptable.

THREE OR MORE PLAYERS

Less easy ...

One possibility : last diminisher method

- First player (of a group of N players) "cuts" a piece that looks fair to that player
- That piece gets examined by the other players, 2 through N, successively. Each of these players can choose to "trim" the piece if they think it is too large for a fair share
- After everybody has inspected it and possibly trimmed it, the piece goes to the last player who chose to diminish it, or to player 1 if nobody did

- The procedure can be repeated with the remainder of the cake for the remaining N-1 players.

Try it out with friends...

The problem with this and many other methods: it is NOT envy-free

~ What's an envy-free solution?

A solution in which, after every player has his/her piece, nobody thinks that someone else is better off.

This is not guaranteed in the above procedure: when the first "piece" is allocated, the player who receives it may be happy with it, but he may change his mind when he sees that later players get much bigger pieces after he has left the division game.

Making fair division envy-free is much harder.

An envy-free division for three players (1960; found independently by John Conway and John Selfridge)

- player 1 cuts cake in three pieces that look equal to that player, and hands over to player 2
- player 2 may, if she wishes, trim the piece that she thinks is largest so that it is equal to the next-largest, in her perception. The trimming T is set aside for the moment.
- player 3 chooses the piece he thinks is largest.
- next player 2 chooses. If 2 did trim in the second step, and if 3 did not take the trimmed piece, then 2 must take the trimmed piece
- 1 gets the remaining piece.

So far they are all happy and there is no envy:

- 3 chose first
- 2 chose and got of the two pieces she considered to be a tie for largest
- 1 got one of the pieces that he cut, and everybody else got (in their eyes) the same or less.

Q: Now, what do you do with the trimming?

Whatever happens with it, player 1 will never envy the player who received the trimmed piece in the first round, because for player 1, trimmed piece + trimming only make up as much as he (1) got in the first round anyway.

let's call the player who got the trimmed piece in the first round Tr (Tr is either 2 or 3) and the other one (of 2 and 3) Untr.

Now Untr will cut the trimming into three equal pieces (from his/her point of view). Then the other players choose.

first Tr, then 1, then Untr takes the last piece of the trimming

Result:
- Tr is happy, and envies no one, because Tr chose first
- 1 does not envy Tr
  1 does not envy Untr because he chose ahead of Untr
- Untr does not envy anyone, because Untr did the cutting

★ No easy way to generalize this to 4 or more players

Remarks
☐ You can also use this to divide up a list of chores!
☐ This can be extended to more complicated problems, such as dividing up an estate.

Dividing up an estate, or property settlement in a divorce

Divorce: Usually only two parties

It is possible to end up with a situation where each party ends up with what they perceive as more than their fair share!

Example. Alice and Bob are divorcing. ☹

They have only two major assets, which need to be divided

First, each of them is asked to allocate points to the two assets, out of a total of 100, according to what they value most.

- Alice is a city person, and places premium value on the small NYC apartment that the couple owns.
- Bob is retired and likes to spend his time fishing; he values their nice shore house much more than the apartment.

	Alice	Bob
shore house	30	70
NYC apt	70	30

In this case, it makes sense to give Alice the Apartment, and Bob the shore house.

In practice, the situation is usually more complicated, with more assets:

Example Bill and Matilda divorce

The point allocation table is not known, of course. Based on the negotations, one can make the following guess:

Asset	Bill	Matilda
Sardinia villa	10	38
Connecticut estate	40	20
Yacht $$$	10	30
NYC plaza apartment	38	10
Cash & jewelry	2	2
	100	100

STEP 1: Give each party the big items that they like most

Bill :	Connecticut estate	40
	NYC plaza apartment	38
		78 points
Matilda:	NYC plaza apartment	38
	Yacht	30
		68 points

STEP 2 : Give the remaining "small" things to the party who has the fewest points, to even out the result as much as possible. In this case, Matilda gets the cash and jewelry, and has now 70 points.

STEP 3: The situation is not even. We need to transfer a bit from Bill to Matilda.

Since Matilda values Connecticut estate over the NYC plaza apartment, while Bill values these two about equally. It makes sense to transfer part of the Connecticut estate. How much?

If we give x% of the Connecticut estate to Matilda, this leaves (100-x)% of the Connecticut estate to Bill

Q : How many points does each of the parties have then?

original points

Bill : $40 \times \left(\frac{100-x}{100}\right) + 38 = \boxed{78} - 0.4x$

Matilda : $20 \times \left(\frac{x}{100}\right) + 70 = \boxed{70} + 0.2x$

original points

To make things even, we require

$$78 - 0.4x = 70 + 0.2x$$
$$8 = 0.6x$$
$$x = \frac{8}{0.6} = 13.3$$

In practice,

- Bill gets the NYC plaza apartment
- Matilda gets the yacht, Sardinia villa, and cash & jewelry
- Bill gets the Connecticut estate 11 months/year
- Matilda got the Connecticut estate 1 month/year

Lecture 21 VORTEX MOTION / FLUID DYNAMICS (based on Jonathan Marshall's) vortex dynamics notes

Equations of motion. To derive these, we'll suppose that every point $\underline{x}$ in the flow domain is occupied at each instant t by a fluid "particle", and then consider the motion of this particle

Material derivative

Suppose $P(\underline{x},t)$ is some property of the fluid (e.g. density, temperature, etc). If x, y, z and t change by small amounts $\delta x, \delta y, \delta z$ and δt, then

$$\delta P = \frac{\partial P}{\partial x}\delta x + \frac{\partial P}{\partial y}\delta y + \frac{\partial P}{\partial z}\delta z + \frac{\partial P}{\partial t}\delta t \qquad (\dagger)$$

If we restrict our attention to the change in P following a fluid particle, which moves with the flow velocity

$$\underline{v}(\underline{x},t) = \big(u(\underline{x},t),\ v(\underline{x},t),\ w(\underline{x},t)\big)$$

then

$$\begin{aligned} \delta x &= u(\underline{x},t)\delta t \\ \delta y &= v(\underline{x},t)\delta t \\ \delta z &= w(\underline{x},t)\delta t \end{aligned}$$

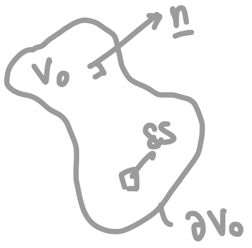
By substituting these into $(\dagger)$ we obtain

$$\begin{aligned} \delta P &= \frac{\partial P}{\partial x}u(\underline{x},t)\delta t + \frac{\partial P}{\partial y}v\delta t + \frac{\partial P}{\partial z}w\delta t + \frac{\partial P}{\partial t}\delta t \\ &= \left(\underline{v}\cdot\nabla P + \frac{\partial P}{\partial t}\right)\delta t \\ &= \delta_{\underline{v}} P \end{aligned}$$

Then we define the <u>material derivative</u> to be

$$\lim_{\delta t\to 0} \delta_{\underline{v}} P = \underbrace{\left[\underline{v}\cdot\nabla + \frac{\partial}{\partial t}\right]}_{\equiv \frac{D}{Dt}} P$$

Conservation of mass Consider a volume V_0 fixed in the fluid



$\rho(\underline{x},t)$ = density of the fluid

The mass $M(t)$ of the fluid in V_0 at time t is given by

$$M(t) = \int_{V_0} \rho(\underline{x},t)\, dV$$

element of volume

$$\left(\rho = \frac{m}{V} \Rightarrow m = \rho V \right)$$

Rate of change of fluid mass in V_0 is

$$\frac{dM}{dt} = \frac{d}{dt}\int_{V_0} \rho(\underline{x},t)dV = \int_{V_0} \frac{\partial \rho}{\partial t}(\underline{x},t)dV \qquad (*)$$

If mass is conserved (no mass created or destroyed) then this rate of change of $M(t)$ must equal the net flux of fluid through ∂V_0. We can write this as

$$-\int_{\partial V_0} \rho \underline{v}\cdot\underline{n}\, dS$$

into fluid surface

But assuming $\underline{v}(\underline{x},t)$ is differentiable in V_0 (which is in keeping with our assumption of mass conservation, then we can apply the divergence theorem (from Multivariable Calculus)

$$\Rightarrow -\int_{\partial V_0} \rho\underline{v}\cdot\underline{n}\,dS = -\int_{V_0} \nabla\cdot(\rho\underline{v})\,dV$$

Thus, comparing with $(*)$ we have

$$\frac{dM}{dt} = \int_{V_0} \frac{\partial \rho}{\partial t}(\underline{x},t)\,dV = 0$$

and this is equal to $-\int_{V_0} \nabla\cdot(\rho\underline{v})\,dV$

So together, we have $\int_{V_0} \frac{\partial \rho}{\partial t}(\underline{x},t)\,dV + \int_{V_0} \nabla\cdot(\rho\underline{v})\,dV = 0$

$$\Rightarrow \int_{V_0} \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \underline{v}) \right] dV = 0$$

But since V_0 is arbitrary this is identically zero iff

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \underline{v}) = 0} \quad (\ddagger)$$

But $\nabla \cdot (\rho \underline{v}) = \rho \nabla \cdot \underline{v} + \underline{v} \cdot (\nabla \rho)$. So $(\ddagger)$ can also be written as

$$\left(\frac{\partial}{\partial t} + \underline{v} \cdot \nabla \right) \rho + \rho \nabla \cdot \underline{v} = 0$$

OR using the material derivative definition:

$$\boxed{\frac{D}{Dt} \rho(\underline{x}, t) + \rho(\underline{x}, t) \nabla \cdot \underline{v}(\underline{x}, t) = 0} \quad (⌂)$$

Here we'll consider incompressible flows $(\nabla \cdot \underline{v} = 0)$. These are ones for which the density of our fluid particles does not change as we move around, i.e. $\frac{D\rho}{Dt} = 0$, or equivalently from (⌂) $\boxed{\nabla \cdot \underline{v}(\underline{x}, t) = 0}$

Streamlines A streamline is a line which at each instant is locally parallel to the velocity field $\underline{v}(\underline{x}, t)$

Then letting $d\underline{x}$ to denote an infinitesimal section of a streamline, $d\underline{x} = k \underline{v}$, where k may depend on $\underline{x}$ and t

So at each point along streamlines we have $d\underline{x} = k \underline{v}$ for k real.

Alternatively $\frac{dx}{u} = \frac{dy}{v} = \frac{dz}{w}$, where $\underline{u} = (u, v, w)$

This system of simultaneous ODEs, together with an initial condition (corresponding to fixing a single point on the streamline), determine the equation of the streamline.

Stream function If we have an incompressible flow in 2D (or 3D with some symmetry e.g. axisymmetric - rotating about some axis in 3D space). then our condition $\nabla \cdot \underline{v} = 0$

$$\Rightarrow \exists \text{ a scalar function } \psi(x, y) \text{ s.t.}$$

$$\boxed{u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x}}$$

Check (Proof not given but converse is easy to check)

$$\nabla \cdot \underline{v} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial^2 \psi}{\partial y \partial x} = 0$$

One can also write the above as $\underline{v} = \nabla \underset{\uparrow}{\times} (\psi \hat{k})$ (Here $\hat{k}$ = unit vector perpendicular to the (x,y)-plane)

cross product

$$\nabla \times (\psi \hat{k}) = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 0 & 0 & \psi(x,y) \end{vmatrix} = \hat{i}\left(\frac{\partial \psi}{\partial y}\right) - \hat{j}\left(\frac{\partial \psi}{\partial x}\right) = \left(\frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x}, 0\right)$$

Now note that $d\psi = \frac{\partial \psi}{\partial x} dx + \frac{\partial \psi}{\partial y} dy + \frac{\partial \psi}{\partial t} dt$ [Recall: $\psi = \psi(\underline{x}, t)$]

Consider $d\psi$ as we move along a streamline fixed at some instant in time.

Time fixed $\Rightarrow dt = 0$

Furthermore, along a streamline $d\underline{x} = k\underline{v} = k\left(\frac{\partial \psi}{\partial y}, -\frac{\partial \psi}{\partial x}\right)$

$$\Rightarrow d\psi = \frac{\partial \psi}{\partial x}\left(k\frac{\partial \psi}{\partial y}\right) + \frac{\partial \psi}{\partial y}\left(-k\frac{\partial \psi}{\partial x}\right) = 0$$

i.e. ψ is constant along each streamline

So $\psi(\underline{x}, t)$ is called the stream-function of the flow.

Examples. ① $(u, v) = (\gamma x, -\gamma y)$ $\gamma \in \mathbb{R} \neq 0$

streamlines $\frac{dx}{u} = \frac{dy}{v} \Rightarrow v\,dx - u\,dy = 0$

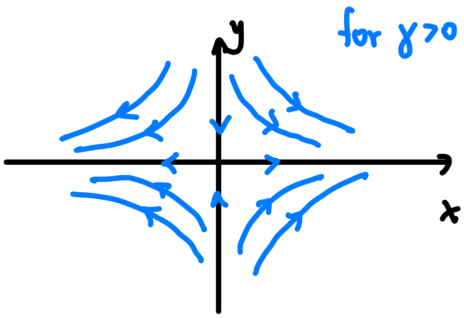
$-\gamma y\,dx - \gamma x\,dy = 0$

$(\div -\gamma)$ $y\,dx + x\,dy = 0$

$d(xy) = 0$

$xy = \text{const}$

$\Rightarrow$ streamlines are hyperbolae



Streamfunction : $\frac{\partial \psi}{\partial y} = u = \gamma x \Rightarrow$ Integrate wrt y: $\psi = \gamma x y + f(x)$

$\frac{\partial \psi}{\partial x} = -v = \gamma y \Rightarrow$ Integrate wrt x: $\psi = \gamma x y + g(y)$

Thus $\psi(x,y) = \gamma xy +$ const

set to 0 without loss of generality

This flow is known as a uniform straining flow

γ is known as the rate of strain

For this case the principal axes of strain are the (x,y) axes.

Lecture 22

② $(u,v) = (-\Omega y, \Omega x)$, $\Omega \in \mathbb{R} \neq 0$

Streamlines: $\frac{dx}{u} = \frac{dy}{v} \Rightarrow v\,dx - u\,dy = 0$

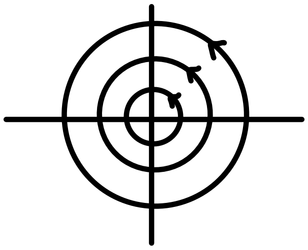
$\Omega x\,dx + \Omega y\,dy = 0$

$x\,dx + y\,dy = 0$

$d(x^2+y^2) = 0$

$x^2+y^2 = \text{const}$

$\Rightarrow$ Streamlines are concentric circles, centered at the origin



For $\Omega > 0$

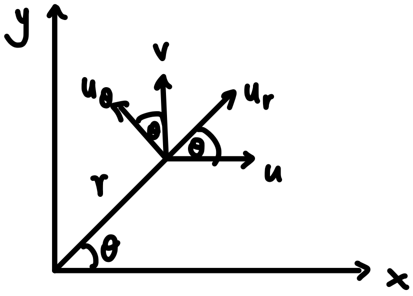
Streamfunction $\frac{\partial \psi}{\partial y} = u = -\Omega y \Rightarrow \psi = -\frac{\Omega y^2}{2} + f(x)$

$\frac{\partial \psi}{\partial x} = -v = \Omega x \Rightarrow \psi = -\frac{\Omega x^2}{2} + g(y)$

$\Rightarrow \psi = -\frac{\Omega}{2}(x^2+y^2) +$ const

0 without loss of generality

It is natural to consider this flow in terms of cylindrical polar coordinates



$$u_r = u\cos\theta + v\sin\theta$$
$$u_\theta = -u\sin\theta + v\cos\theta$$

But also $u = -\Omega y = -\Omega r\sin\theta$

$v = \Omega x = \Omega r\cos\theta$

$$\Rightarrow u_r = -\Omega r\sin\theta\cos\theta + \Omega r\cos\theta\sin\theta = 0$$
$$u_\theta = \Omega r\sin^2\theta + \Omega r\cos^2\theta = \Omega r(\sin^2\theta + \cos^2\theta) = \Omega r$$

Next, angular velocity is defined as $\frac{u_\theta}{r}$. In this case, this is Ω. This is independent of position. Hence the fluid moves like a solid-body. For this reason, this flow is known as solid-body rotation.

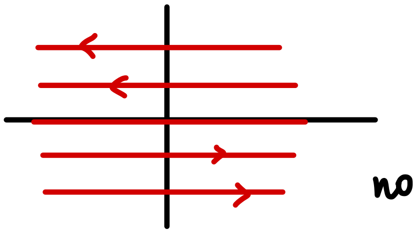
Vorticity The vorticity field $\underline{\omega}$ of a flow $\underline{v}(\underline{x},t)$ is defined by $\boxed{\underline{\omega} = \nabla \times \underline{v}}$.

In 2D $$\underline{\omega} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u(x,y) & v(x,y) & 0 \end{vmatrix} = \left(0, 0, \underbrace{\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}}_{\equiv\ \omega(x,t)}\right)$$

If $\underline{\omega} \equiv 0$ then the flow is said to be irrotational.

A vortex line is a line which defined at some instant in time, which is locally parallel to the vorticity field at each point along it.

③ $(u,v) = (-2\Omega y, 0)$, $\Omega \in \mathbb{R} \neq 0$ incompressible flow



(Shape of streamlines is horizontal lines)

note: $|u|$ increases as $|y|$ increases.

Alternatively, look at the streamfunction

$$\frac{\partial \psi}{\partial y} = u = -2\Omega y$$

$$\frac{\partial \psi}{\partial x} = -v = 0$$

$$\Rightarrow \psi = -\Omega y^2$$ lines of constant ψ are streamlines.

This is known as a shear flow

$$\omega(x,y) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 2\Omega$$

Like solid body rotation, this has constant vorticity everywhere. However, these two flows look very different. But we can write

$$\underset{\text{shear flow}}{\underset{\uparrow}{(-2\Omega y, 0)}} = \underset{\substack{\text{solid body}\\ \text{rotation}\\ \text{(s.b.r.)}}}{\underset{\uparrow}{(-\Omega y, \Omega x)}} + \underset{\hat{\underline{v}}\ ?}{\underset{\uparrow}{(-\Omega y, -\Omega x)}}$$

– What sort of flow is $\hat{\underline{v}}$?

Clearly incompressible $\nabla \cdot \hat{\underline{v}} = 0$

This has vorticity $\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = -\Omega + \Omega = 0$

i.e. $\hat{\underline{v}}$ is irrotational.

Note that the vorticity of shear flow $= 2\Omega$, and the vorticity of solid body rotation $= 2\Omega$

$\therefore$ expect $\hat{\underline{v}}$ will have vorticity $= 0$

Streamfunction:

$$\frac{\partial \psi}{\partial y} = u = -\Omega y \Rightarrow \psi = -\frac{\Omega}{2}y^2 + f(x)$$

$$\left(\frac{\partial \psi}{\partial x} = -v = \Omega x\right) \qquad \frac{\partial \psi}{\partial x} = f'(x) = \Omega x$$

(compare with term in brackets to infer $f'(x) = \Omega x$)

Thus $$\psi = \frac{\Omega x^2}{2} + \frac{\Omega y^2}{2} = \frac{\Omega}{2}(x^2 - y^2)$$

This is to be expected as $$\psi_{\text{shear}} = -\Omega y^2$$

$$\psi_{\text{s.b.r}} = \frac{\Omega}{2}(x^2 + y^2)$$

Linear combination of stream functions

$$\psi_{shear} = \psi_{s.b.r} + \psi_{\hat{\underline{v}}}$$

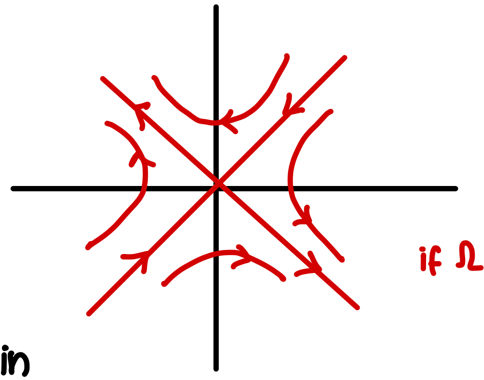
$$\Rightarrow \quad -\Omega y^2 = -\frac{\Omega}{2}(x^2+y^2) + \psi_{\hat{\underline{v}}}$$

$$\Rightarrow \quad \psi_{\hat{\underline{v}}} = \frac{\Omega}{2}(x^2-y^2)$$

Thus, streamlines are given by $x^2-y^2=\text{const}$

i.e. they are hyperbolae

In particular $x^2-y^2=0 \Rightarrow y=\pm x$



$(u,v)=(-\Omega y,-\Omega x)$

if $y<0 \Rightarrow u>0$
if $x>0 \Rightarrow v<0$

$\hat{\underline{v}}$ is a straining flow, but with principal axes of strain along $y=\pm x$.

Thus, this shear flow can be considered as the sum of a solid body rotation and a straining flow. This is in fact true locally for every incompressible flow.

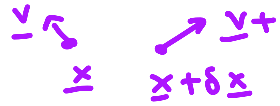
Lecture 22

Show this as follows:

LOCAL ANALYSIS

Consider a 2D incompressible flow $\underline{v} = (u(x,y), v(x,y))$

Consider the velocity field relative to some point $\underline{x}$ in the flow domain i.e.



$\underline{v}(\underline{x}+\delta\underline{x}) - \underline{v}(\underline{x}) = \delta\underline{v}$, say where $\delta\underline{x} = (\delta x, \delta y)$

Note $u(x+\delta x, y+\delta y) = \underset{\bigstar}{u(x,y)} + \frac{\partial u}{\partial x}\delta x + \frac{\partial u}{\partial y}\delta y + O(\delta x^2), O(\delta y^2)$

similarly for v. Taylor series expansion (multivariate expansion)

$$\Rightarrow \boxed{\delta\underline{v}} = \begin{pmatrix} \frac{\partial u}{\partial x}\delta x + \frac{\partial u}{\partial y}\delta y \\ \frac{\partial v}{\partial x}\delta x + \frac{\partial v}{\partial y}\delta y \end{pmatrix} = \boxed{A\delta\underline{x}}$$ where $A = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$ and $\delta\underline{x} = \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}$

column vector

Note that the $u(x,y)$ and $v(x,y)$ as in $\bigstar$ do not appear because we subtract $\underline{v}(\underline{x})$ from $\underline{v}(\underline{x}+\delta\underline{x})$

We can write $\boxed{A = E + F}$ where $F = \begin{pmatrix} 0 & -\frac{\omega}{2} \\ \frac{\omega}{2} & 0 \end{pmatrix}$ where $\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ (scalar vorticity field)

we defined this on pg 101

and $E = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{1}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right) & \frac{\partial v}{\partial y} \end{pmatrix}$

Now note that $\delta \underline{v} = A\delta\underline{x} = (E+F)\delta\underline{x} = E\delta\underline{x} + F\delta\underline{x}$

But $F\delta\underline{x} = \begin{pmatrix} 0 & -\frac{\omega}{2} \\ \frac{\omega}{2} & 0 \end{pmatrix}\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = \frac{\omega}{2}\begin{pmatrix} -\delta y \\ \delta x \end{pmatrix}$

This corresponds to a solid body rotation about $\underline{x}$ with angular velocity $\frac{\omega}{2}$

$\underline{Q}$ What about E?

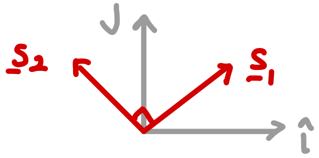
Since E is real and symmetric, it has real eigenvalues λ_1 and λ_2, say (not necessarily distinct). Also, there exists an orthonormal basis of $\mathbb{R}^2$ consisting of eigenvectors $\underline{s}_1$ and $\underline{s}_2$ (these are column vectors) of E (where $E\underline{s}_j = \lambda_j \underline{s}_j$ for $j = 1, 2$).

And we can diagonalize E to write $E = SMS^T$ where $M = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, $S = (\underline{s}_1, \underline{s}_2)$

So $E\delta\underline{x} = SMS^T\delta\underline{x}$.

Let $\delta\underline{x}' = S^T\delta\underline{x} = \begin{pmatrix} \underline{s}_1 \cdot \delta\underline{x} \\ \underline{s}_2 \cdot \delta\underline{x} \end{pmatrix}$

shift of coordinates to a different frame of reference



$\underline{s}_j \cdot \delta\underline{x}$ is just the component of $\delta\underline{x}$ in the direction of $\underline{s}_j$.

To get a qualitative idea of the nature of the flow corresponding to E, it is enough to consider $M\delta\underline{x}'$, since S simply shifts this back to our original basis $\{\hat{i}, \hat{j}\}$.

Note that if $E = SMS^T$ then from linear algebra we know that

$$\begin{aligned} \text{trace } M &= \text{trace } E \\ &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \\ &= \nabla \cdot \underline{v} \\ &= 0 \end{aligned}$$

since the flow is assumed to be incompressible.

But recall that $m = \begin{pmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{pmatrix}$ so $\text{trace}(m) = \lambda_1 + \lambda_2 = 0$

$\Rightarrow \lambda_1 = -\lambda_2 = \gamma$ say

$$M\delta\underline{x}' = \begin{pmatrix}\gamma & 0 \\ 0 & -\gamma\end{pmatrix}\begin{pmatrix}\delta x' \\ \delta y'\end{pmatrix} = \gamma(\delta x', -\delta y')$$

Observe that this corresponds to a uniform straining flow with principal axes of strain in the directions of $\underline{s}_1$ & $\underline{s}_2$ and straining rate γ.

⚠ Important: Vorticity corresponds to LOCAL not global rotation of a fluid

To highlight this, consider the following example.

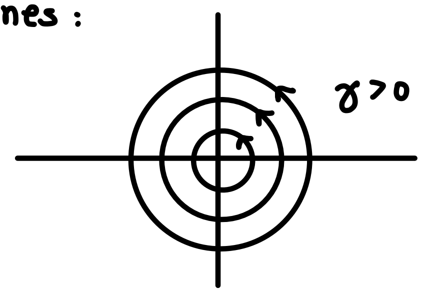
Example: Consider the flow $(u_r, u_\theta) = \left(0, \frac{\gamma}{2\pi r}\right)$ $\gamma \in \mathbb{R}$

- movement in azimuthal direction
- no movement in radial direction.

One can check that this is an incompressible flow: $\nabla \cdot \underline{v} = \frac{1}{r}\left(\frac{\partial(r u_r)}{\partial r} + \frac{\partial u_\theta}{\partial \theta}\right) = 0$

As for solid body rotation, this flow is purely in the azimuthal direction.

Streamlines:



So globally the fluid rotates about the origin. However

$$\underline{\omega} = \nabla \times \underline{v} = \begin{vmatrix} \underline{e}_r & r\underline{e}_\theta & \underline{e}_z \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial z} \\ u_r & r u_\theta & u_z \end{vmatrix} = 0$$

2D space embedded in 3D space for vorticity field (even for 2D $\underline{w}$ is in $\underset{\sim}{k}$-direction)

where $u_r = 0$

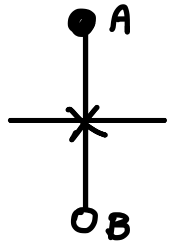
$u_\theta = \frac{\gamma}{2\pi r}$

if $r \neq 0$ (at $r=0$ the flow is singular)

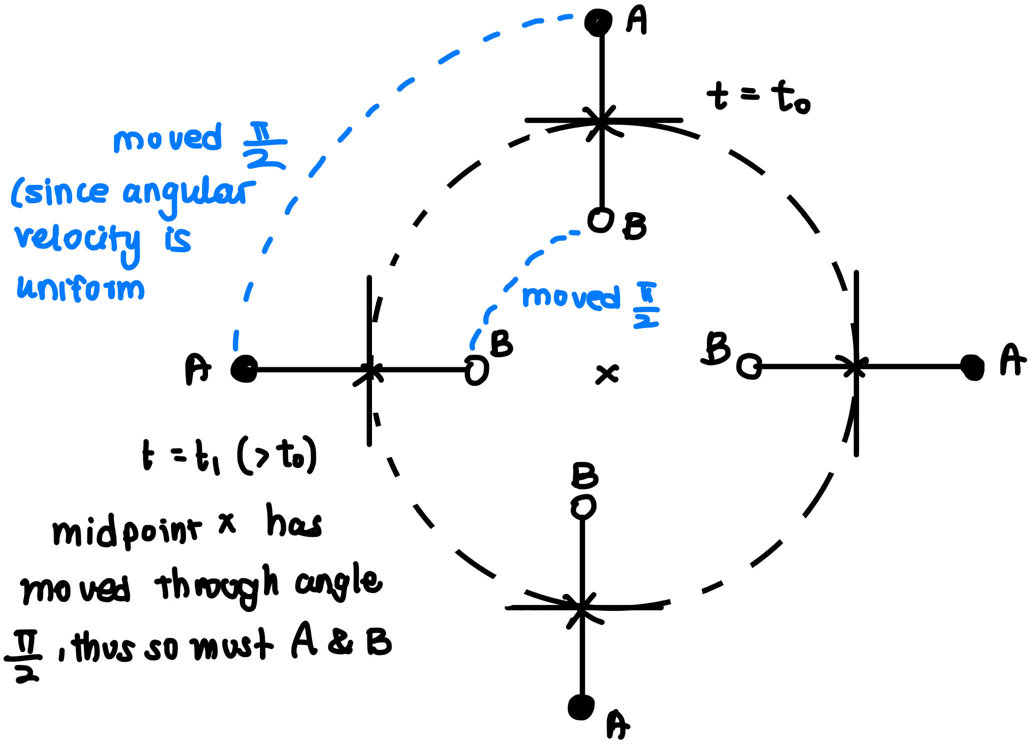
So there is no local rotation about non-zero points (i.e. origin).

One may examine the difference between this singular flow and solid body rotation as follows. (based on Acheson's book: Elementary fluid dynamics)

Consider a vorticity meter



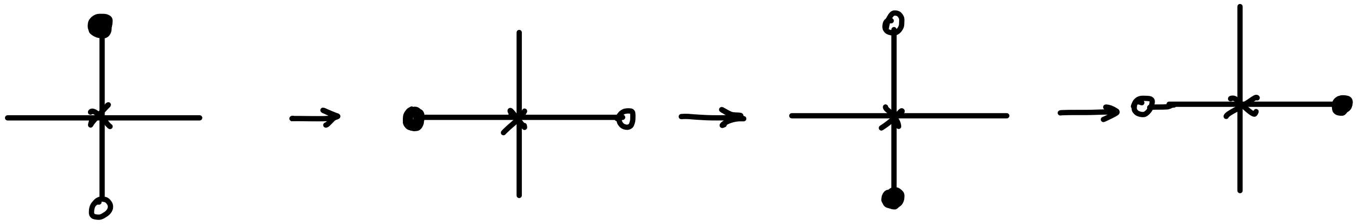
For solid body rotation:



angular velocity is uniform (i.e. the same at all points)

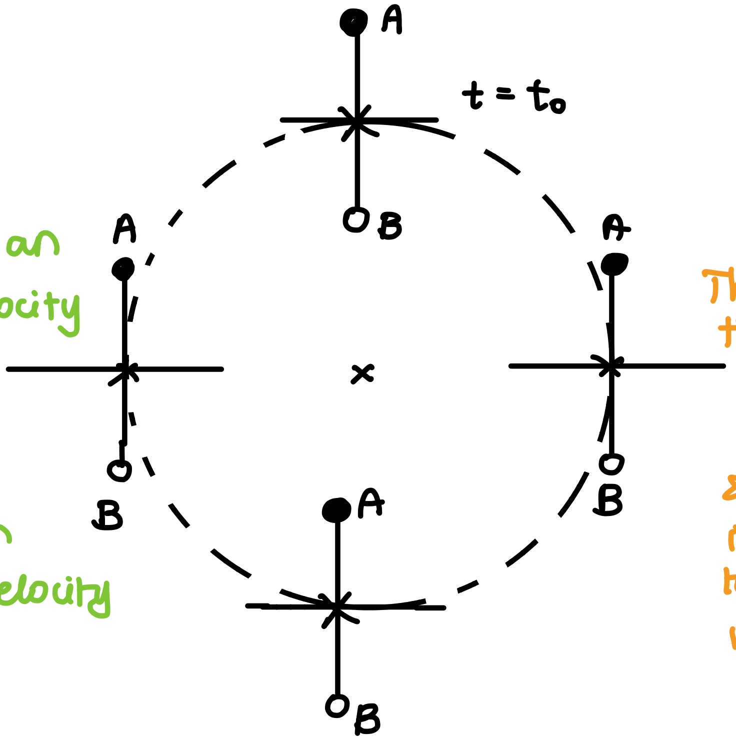
midpoint x has moved through angle $\frac{\pi}{2}$, thus so must A & B

Considering motion relative to midpoint x, we observe



i.e. local rotation at non-zero points.

For our singular flow $(u_r, u_\theta) = \left(0, \frac{\gamma}{2\pi r}\right)$, however the angular velocity is not uniform; it decreases as r increases. In fact it varies in precisely the right way so that one observes the following:



moved through an angle $< \frac{\pi}{2}$ since velocity is smaller

moved through an angle $> \frac{\pi}{2}$ since velocity is larger

$\downarrow u_\theta = \frac{\gamma}{2\pi r \uparrow}$

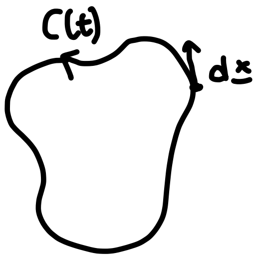
The distance of A from the origin is bigger than the distance of the midpoint from the origin & so the velocity of A is smaller. B is closer to the origin so its velocity is bigger

i.e. there is no local rotation about the midpoint x (or in fact any other point not at the origin) ⇒ zero vorticity.

This singular flow is in fact called a point vortex flow.

As a measure of global rotation of a fluid flow we introduce the following

Circulation Let C(t) be a closed contour in the flow domain each point along which moves with the local velocity field.



The circulation Γ(t) around C(t) is defined to be $\boxed{\Gamma(t) = \oint_{C(t)} \underset{\sim}{v} \cdot d\underset{\sim}{x}}$

where $\underset{\sim}{v}$ is the velocity field, $d\underset{\sim}{x}$ is a vector of infinitesimal length, tangential to C(t). and we integrate round C(t) with its interior on our left.

$\Gamma(t)$ can be interpreted as a measure of the flow around $C(t)$. Alternatively, applying Stokes theorem,

$$\Gamma(t) = \iint_{S(t)} (\underbrace{\nabla\times\underline{v}}_{\overset{||}{\underline{\omega}}\ :\text{vorticity field}})\cdot\underline{n}\,dS$$

where $S(t)$ is any surface spanning $C(t)$ and $\underline{n}$ is the unit normal (whose direction is given by the "right-hand rule". (Note that here, in order to apply Stokes theorem we've assumed $\underline{v}$ to be non-singular in S).

So $\Gamma(t)$ can also be thought of as the flux of vorticity through S.

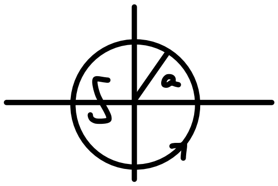
In <u>2D</u> we have $\underline{n} = (0,0,1)$, $\underline{\omega} = (0,0,\omega)$ $\Rightarrow$ $\boxed{\Gamma(t) = \iint_{S(t)} \omega\, dS}$

$\uparrow$ unit vector in z-direction

Examples ① Solid body rotation

$$(u,v) = (-\Omega y, \Omega x)$$

Pick C to be the circle $x^2+y^2=a^2$



$$\Gamma = \oint_C \underline{v}\cdot d\underline{x}$$
$$= \Omega\oint_C (-y\,dx + x\,dy)$$

For (x,y) on C $\begin{Bmatrix} x = a\cos\theta \\ y = a\sin\theta \end{Bmatrix}$ $\rightarrow$ $\begin{aligned} dx &= -a\sin\theta\, d\theta \\ dy &= a\cos\theta\, d\theta \end{aligned}$

Therefore, $\Gamma = \Omega\int_0^{2\pi} a^2(\underbrace{\sin^2\theta + \cos^2\theta}_{=1})\,d\theta = 2\pi\Omega a^2$

const. circulation around contour C

Alternatively, recalling $\omega = \dfrac{\partial v}{\partial x} - \dfrac{\partial u}{\partial y} = 2\Omega$

$\Gamma = \iint_S 2\Omega\, dS = 2\pi\Omega a^2$ as before (area of circle $S = \pi a^2$)

② Point vortex

$(u_r, u_\theta) = \left(0, \frac{\gamma}{2\pi r}\right)$ ← irrotational everywhere except origin (i.e. cirulation $=0$)

$$\Gamma = \oint_C \underline{v} \cdot d\underline{x} \qquad d\underline{x} = (dr, r d\theta)$$

$$\underline{v} \cdot d\underline{x} = \begin{pmatrix} 0 \\ \frac{\gamma}{2\pi r} \end{pmatrix} \cdot \begin{pmatrix} dr \\ r d\theta \end{pmatrix} = \frac{\gamma}{2\pi} d\theta$$

$$= \oint_C \frac{\gamma}{2\pi} d\theta$$

$$= 2\pi \frac{\gamma}{2\pi}$$

$$= \gamma \neq 0$$

This is non-zero due to the singularity of the flow at the origin. In fact, the vorticity distribution for this flow is

$$\omega(x, y) = \delta(x, y)$$

where $\delta(x, y)$ is the delta-function. $\delta(x-x', y-y') = 0$ for $(x, y) \neq (x', y')$

and $\iint_S \delta(x-x', y-y')\, dS = 1$ if $(x', y') \in S$

$$\Gamma = \iint \omega\, dS$$

Lectures 23 + 24 In the next few classes we'll have an introduction to $\boxed{\text{MATLAB.}}$ Use this as an opportunity to practice writing code on your own as it's the best way to learn! This will also be useful for your final projects. (MATLAB Crash Course, Univ. of Oxford)

Useful references:
- D.J. Hingham and N.J. Higham, MATLAB Guide, SIAM, 2005
- T A. Driscoll, Learning MATLAB, SIAM, 2009
- C.B. Moler, Numerical Computing with MATLAB and Experiments with MATLAB (freely available online. http://www.mathworks.com/moler/)
- Online MATLAB courses: https://matlabacademy.mathworks.com/
- MATLAB Cody. https://www.mathworks.co.uk/matlabcentral/cody/

Tentative timetable

Day 1: Introduction

Theory 1: Basic operations with the command window

Practical 1

Theory 2: Scripts, logic, control structures & anonymous functions

Practical 2

Day 2: Theory 3: Cell arrays, functions, and programming

Practical 3

Theory 4: Graphics

Questions:

- How many of you used MATLAB before?
- How many have coded in another language?

```
%% Theory1:
% MATLAB Crash Course: Basic operations with the command window.
%
% Originally written by Nick Hale, Oct. 2013.
% Extended by Asgeir Birkisson, Oct. 2014, 2015.
% Modified by Behnam Hashemi and Hadrien Montanelli, Sep. 2016.

%% First steps
5 + 10
3 - 2
3*2
3/2
3^2
exp(3)
sqrt(9)
factorial(5)
sin(3)
sin(pi)
sind(90)

%% Get help
help sin
doc sin
```

```
%% Initialize vectors
a = [1 3 5]      % Row vector
a = [1, 3, 5]    % The same
size(a)          % Size of a
length(a)        % max of the above
a = [1 ; 3 ; 5]  % Column vector
size(a)          % Size of a
a = [1+1i 3 5]   % Column vector with complex entries
a = [1+1i 3 5].' % .' gives the transpose
a = [1+1i 3 5]'  % ' gives the conjugate transpose

%% Simple commands
clc             % clear command window
a
max(a)          % Maximum value
min(a)          % Minimum value
sum(a)          % Sum of entries
mean(a)         % Average value

%% Addition and multiplication
b = [2 6 10]';  % Another column vector
c = a + b
d = 4*a
e = 3*a + 5*b;

%% Modifying a vector
a
a(2) = 11       % Modify second entry
a = [a; 4]      % Add an entry at the end
a = [7; a]      % Add an entry at the start
a(3) = []       % Remove the third entry

%% Vector syntax
1:100
1:5:101
10:-2:0
linspace(0, 1, 51)

%% Initialise a matrix
A = [1 8; 5 2]  % 2x2 matrix
A'              % (Conjugate) Transpose of the matrix
size(A)
length(A)

%% Simple commands -- acting columnwise
max(A)
min(A)
sum(A)
mean(A)

%% Simple commands -- acting rowwise
% Notice extra arguments to the function
max(A, [], 2)
min(A, [], 2)
sum(A, 2)
mean(A, 2)

%% Addition and multiplication
B = [4 5; 9 3];   % Another 2x2 matrix
C = 3*A + B
```

```
%% Matrix syntax
A(1, 2)
A(:, 2)
A(1, :)
D = diag(A)        % Diagonal elements
det(A)

%% Useful commands
A = rand(3, 3)  % matrix with random elements
A = rand(3)     % the same
O = ones(3)     % matrix with ones
Z = zeros(3)    % matrix with zeros

%% Factorizations
A = rand(5)
[V, D] = eig(A)    % Eigenvectors and eigenvalues
[L, U, P] = lu(A)  % LU decomposition
[Q, R] = qr(A)     % QR factorisation
Q*Q'

%% Solve a linear system -- Ax = b

% Solve
%     x1  + 2*x2 = 1
%    5*x1 + 8*x2 = 2
A = [1 2; 5 8];
b = [1 2]';
x = A\b             % Use backslash for solving
x = inv(A)*b;       % This is not good -- numerical instabilities

% Solve with random coefficients and right-hand side:
A = rand(2, 2);
b = rand(2, 1);
x = A\b

%% Formats
pi
format long
pi
% Format does NOT affect how Matlab computations are done, just the display
format short
a = sqrt(2)
format long
b = sqrt(2)
a - b

% Get rid of extra linespaces
format compact
a - b

% Reintroduce the extra linespaces
format loose
a - b

%% Basic plotting

x = linspace(-1, 1, 100);
plot(x, sin(4*pi*x))
%%
hold on
plot(x, exp(cos(10*x)), 'r')
hold off
```

For the last two lectures we will look at the basic principles of CONTROL THEORY
(based on notes by Hartmann, NYU Berlin)

Lecture 25

We'll start with an example that considers fishering management.

The question we'll try to answer is:
Is there an optimal harvesting strategy that maximizes the sustainable catch or that maximizes the profit on a time-horizon of several years?

Assumption: No interaction between different species

Based on the logistic population model for a single species.

We introduce the following functions:

$x(\cdot) \in \mathbb{R}$ $x(t)$ = fish population at time t
$b(\cdot) \in \mathbb{R}$ $b(t)$ = number of boats operating at time t
$h(\cdot) \in \mathbb{R}$ $h(t)$ = harvesting rate at time t

Note: for simplicity we assume that all functions take real values, even though the number of boats will be an integer number

HARVESTING STRATEGY: Controlling the number of boats used for fishing

Call b the control variable even though it is a piecewise defined function $b: [0,\infty) \to \mathbb{R}$

We consider the following parameters

$C_B > 0$: overhead cost per boat and unit of time
n : number of fishermen per boat
w : fishermen's salary per unit of time
p : market price of one unit of fish

The boundary conditions and available parameters determine what a good harvesting strategy is.

e.g Maximizing the sustainable catch is different from maximizing the long-term profit, which may be different from maximizing the short-term profit.

★ The answer depends on the question ★

SETTING UP THE MODEL

Relate the harvest rate h with the number of fish x and the number of boats b

NOTE The $\boxed{\text{static relation}}$ between these variables is called a Constitutive relation.

This is different from the dynamic relation between different species in a predator-prey model.

e.g. Hooke's law is a constitutive relation (kinematic relation between the force exerted by a spring and its extension).

Newton's law expresses a dynamical relation between force and acceleration.

Here we assume that the harvesting rate is proportional to both the number of fish and the number of boats, i.e. we assume the following relation

$$\boxed{h(t) = q\, x(t)\, b(t)}$$

Constitutive relation

Where $q > 0$ is a constant of proportionality that depends on the efficacy of the fishing boats.

The harvesting rate is the rate by which the growth rate of a fish population is reduced as an effect of fishing.

We assume that the fish population evolves acording to the logistic equation

$$\boxed{\frac{dx}{dt} = \gamma x\left(1 - \frac{x}{K}\right) - h, \qquad x(0) = x_0 > 0} \quad (+)$$

$\gamma > 0$: initial growth rate of the population when x is small

$K > 0$: Capacity of the ecosystem without fishing

Maximizing any given objective, such as sustainable catch or profit under the constraint that the fish population evolves according to the dynamics given by (†) is not possible without further specifying what the admissible controls $b(\cdot)$ are.

Assume that the only admissible strategies are of the form

$$b : [0,\infty) \to \mathbb{R}, \qquad b(t) = \begin{cases} 0 & t \leq t^* \\ b_0 & t > t^* \end{cases}$$

with the two adjustable, but a priori unknown parameters $t^* \geq 0$, $b_0 > 0$

Thus, our harvesting strategy can be controlled by choosing the right time t^* at which fishing is started and the corresponding number b_0 of boats.

Resulting logistic model is a switched ODE of the form:

$$\frac{dx}{dt} = \begin{cases} \gamma x\left(1 - \frac{x}{K}\right), & t \leq t^* \\ \gamma x\left(1 - \frac{q b_0}{\gamma} - \frac{x}{K}\right), & t > t^* \end{cases}$$

Recall that we had the constitutive relation $h(t) = q x(t) b(t)$ and so at $t > t^*$ we have $h(t) = q x(t) \boxed{b_0}$. Thus at $t > t^*$ $\frac{dx}{dt} = \gamma x\left(1 - \frac{x}{K}\right) - h = \gamma x\left(1 - \frac{x}{K}\right) - q x b_0$

$$= \gamma x\left(1 - \frac{q}{\gamma} b_0 - \frac{x}{K}\right) \quad ✓$$

Maximizing the sustainable catch

Suppose we want to choose b_0 so that the average long-term catch is maximized

→ We must not overfish, otherwise the fish population goes extinct and hence the long-term catch is zero.

For the average long-term catch it does not matter how t^* is chosen, so we can set it to zero and ignore it from now on.

We identify the sustainable population under fishing with the asymptotically stable equilibrium of the system for $b_0 > 0$.

Asymptotic stability is essential for the long-term catch because it is this that guarantees that under small perturbations the equilibrium is robust.

In other words, the population returns to its equilibrium size after a small perturbation that may be, e.g. due to fluctuating environmental conditions.

If one is fishing at an unstable equilibrium instead the fluctuations may cause the population to drift away from its equilibrium and eventually go extinct.

↙ recall this is the initial growth rate of the population when x is small

Lemma. Let $\gamma > q b_0$. Then $x^* = \left(1 - \frac{q b_0}{\gamma}\right) K$ is the unique stable equilibrium.

At eqm, $\frac{dx}{dt} = 0$:

$$\frac{dx}{dt} = \gamma x \left(1 - \frac{q b_0}{\gamma} - \frac{x}{K}\right) = 0 \Rightarrow x^* = 0 \quad \text{OR} \quad 1 - \frac{q b_0}{\gamma} - \frac{x^*}{K} = 0$$

$$x^* = K\left(1 - \frac{q b_0}{\gamma}\right)$$

Note The assumption $\boxed{\gamma > q b_0}$ makes sure that the fish population, growing with rate γ when sufficiently far away from the capacity limit, is not eaten up by the fishing. For $\gamma < q b_0$ the single stable equilibrium is $x^* = 0$.

The solution of the logistic equation for $\underset{\sim}{b_0=0}$ is found using separation of variables

$$\frac{dx}{dt} = \gamma x\left(1-\frac{x}{K}\right)$$

$$\int \frac{dx}{x\left(1-\frac{x}{K}\right)} = \int \gamma \, dt$$

Express the left-hand side integrand as a partial fraction

$$\frac{1}{x\left(1-\frac{x}{K}\right)} = \frac{A}{x} + \frac{B}{\left(1-\frac{x}{K}\right)}$$

$$A\left(1-\frac{x}{K}\right) + Bx = 1$$

let $x=0$: $A=1$

$x=K$. $B=\frac{1}{K}$

Thus $\frac{1}{x\left(1-\frac{x}{K}\right)} = \frac{1}{x} + \frac{1}{K\left(1-\frac{x}{K}\right)}$. Going back to integration using separation of variables, we have

$$\int\left(\frac{1}{x}+\frac{1}{K\left(1-\frac{x}{K}\right)}\right)dx = \int \gamma \, dt$$

$$\ln|x| - \ln|K-x| = \gamma t + C$$

$$\ln\left|\frac{x}{K-x}\right| = \gamma t + C$$

$$\frac{x}{K-x} = Ae^{\gamma t}$$

$$x = KAe^{\gamma t} - Axe^{\gamma t}$$

$$x\left(1+Ae^{\gamma t}\right) = KAe^{\gamma t}$$

$$x(t) = \frac{KAe^{\gamma t}}{1+Ae^{\gamma t}}$$

where γ, K are parameters of the model but A comes from the integration constant and can thus be determined from the initial condition $x(0) = x_0$.

$$x(0) = \frac{KA}{1+A} = x_0$$

$$KA = x_0 + Ax_0$$

$$A(K - x_0) = x_0$$

$$A = \frac{x_0}{K - x_0}$$

Thus, the solution to the logistic equation with $b_0 = 0$ is

$$x(t) = \frac{K\frac{x_0}{K-x_0}e^{\gamma t}}{1 + \frac{x_0}{K-x_0}e^{\gamma t}} = \frac{Kx_0e^{\gamma t}}{K - x_0 + x_0e^{\gamma t}} = \frac{Kx_0e^{\gamma t}}{K + x_0(e^{\gamma t} - 1)}$$

So, the solution to the logistic equation satisfies

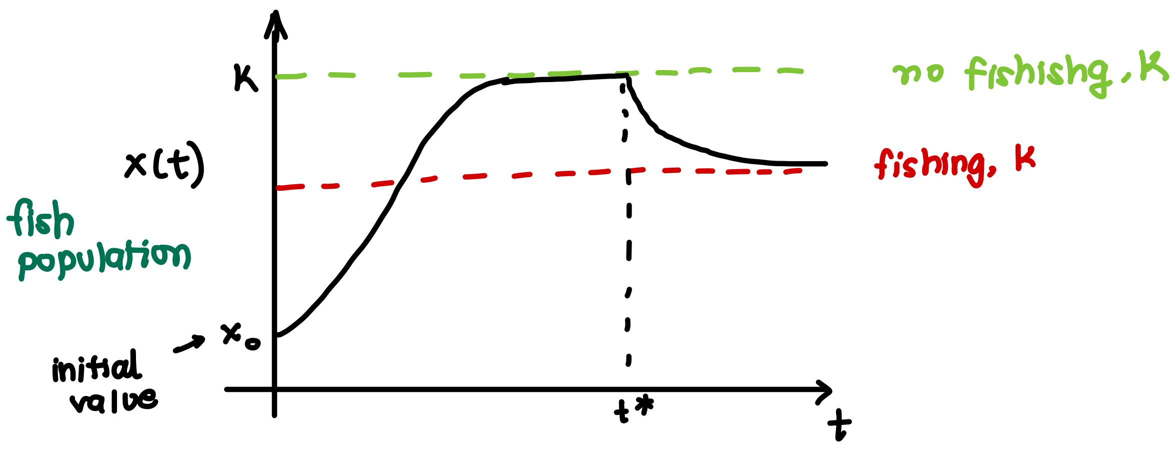
capacity of ecosystem <u>without</u> fishing

$$\lim_{t \to \infty} x(t) = \frac{Kx_0}{x_0} = K$$

(with $b_0 = 0$)

The fishing reduces the capacity of the ecosystem by a factor $1 - \frac{qb_0}{\gamma}$.

A solution of this model would look as follows



lecture 26

We now define the average long-term catch as

$$J_0(b_0) = \lim_{T\to\infty} \int_0^T h(t)\,dt$$

where the expression for the associated sustainable catch rate follows from $h(t) = q\,x(t)\,b(t)$

↑ the constitutive relation

and it takes the form $h(t) = q\,x^* b_0$

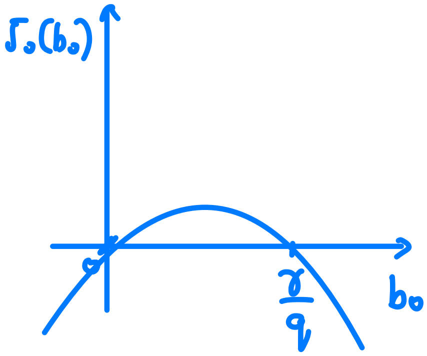
By the lemma, since the asymptotically stable eqm is

$$x^* = K\left(1 - \frac{q b_0}{\gamma}\right)$$

we have that $J_0(b_0) = qK\left(1 - \frac{q b_0}{\gamma}\right) b_0$

The function $J_0(\cdot)$ is strictly concave, which implies that it has a unique maximum

input is b_0, so



The maximizer $b_0^* = \text{argmax}\, J_0(b_0)$ is given by $\boxed{b_0^* = \frac{\gamma}{2q}}$, which rounded to the nearest integer gives the optimal number of fishing boats.

The corresponding optimal sustainable catch is $x^* = K\left(1 - \frac{q b_0^*}{\gamma}\right) = K\left(1 - \frac{\cancel{q}}{\cancel{\gamma}}\left(\frac{\cancel{\gamma}}{2\cancel{q}}\right)\right) = \frac{K}{2}$

e.g. nets used, etc ↓

We see that the maximum sustainable catch is independent of the efficacy q which seems counterintuitive. However if we realize that b_0^* is inversely proportional to q, it makes sense since it makes the optimal harvesting rate independent of q.

A lower efficiency requires more boats and vice versa.
With too many boats the fish population is depleted too much which results in lower catch. The same happens when too few boats are at work, which conserves the fish population, but is suboptimal in terms of the catch.

Optimal control

We just saw that the function J_0 is symmetric about its maximum so if the optimal number of boats was eg $b_0 = 4.6$, the sustainable catch with $b_0 = 5$ boats would be slightly higher than with $b_0 = 4$.
However, if we take into acount that fishing boats are costly, $b_0 = 4$ will be probably be the more reasonable choice.

Objective functional. maximizing profit

We now want to maximize profit rather than catch. So we need to take into account

- costs of maintaining a fishing fleet,
- the market place of fish, etc.

Definition: $\boxed{\text{profit} = \text{revenue} - \text{cost}}$

Profit rate = revenue rate − rate of total costs.

Using that [revenue is the catch times the market price of fish] and that the [total cost is the sum of the overhead costs and the salaries of the fishermen], i.e.

$$\boxed{P(t) = p\,h(t) - (c_B + nw)\,b(t)}$$ profit rate

Recall, last time we defined the following parameters:

$c_B > 0$: overhead cost per boat and unit of time

n : number of fishermen per boat

w : fishermen's salary per unit of time

p : market price of one unit of fish

The total profit until time $t=T$ is then obtained by integrating the profit rate from $t=0$ to $t=T$. To simplify this, we assume that $T=\infty$ and we discount the future profit with a constant discount rate $\delta>0$.

Together with the constitutive relation $h(t)=q\,x(t)\,b(t)$, the overall profit as a function of b becomes

$$\begin{aligned} J(b) &= \int_0^\infty [p\,h(t) - (c_B+nw)\,b(t)]\,e^{-\delta t}\,dt \\ &= \int_0^\infty [p\,q\,x(t)\,b(t) - (c_B+nw)\,b(t)]\,e^{-\delta t}\,dt \\ &= \int_0^\infty b(t)\,[p\,q\,x(t) - (c_B+nw)]\,e^{-\delta t}\,dt \\ &= \int_0^\infty b(t)\,[p\,q\,x(t) - c]\,e^{-\delta t}\,dt \qquad \text{total profit} \end{aligned}$$

where we used $\boxed{c := c_B + nw}$. The discount factor δ accounts for inflation, interest rates or the fact that future rewards are less profitable than immediate rewards. It also ensures that J is finite for our choice of admissible control variables $b(\cdot)$.

Extremum principle We want to maximize the overall profit

$$J(b) = \int_0^\infty b(t)\,[p\,q\,x(t) - c]\,e^{-\delta t}\,dt$$

over all admissible harvesting strategies, i.e. over the switching time t^* and the number of boats b_0.

Since the population $x(t)$ depends on this choice, our optimal harvesting problem is of the form of a maximization problem with a constraint:

$$\max_{b(\cdot)} J(b) \qquad (\dagger)$$

over the set of admissible control strategies $b:[0,\infty)\to\mathbb{R}$, $b(t) = \begin{cases} 0 & t\le t^* \\ b_0 & t>t^* \end{cases}$

and subject to $$\frac{dx}{dt} = \begin{cases} r x\left(1-\frac{x}{K}\right) & t\le t^* \\ r x\left(1-\frac{q b_0}{r}-\frac{x}{K}\right) & t>t^* \end{cases} \qquad (\ddagger)$$

Generally, problems of this form can be solved by the method of Lagrange multipliers or by eliminating the constraint.

A good reference book for this is

Optimal control: Basics and beyond Peter Whittle, 1996

since $b(t)=0$ for $t\le t^*$

$$\text{Note that } J(b)=\int_0^{t^*} \cancelto{0}{b(t)\left[pqx(t)-c\right]} e^{-\delta t}dt + \int_{t^*}^{\infty} b(t)\left[pqx(t)-c\right]e^{-\delta t}dt$$

$$=\int_{t^*}^{\infty} b_0\left[pqx(t)-c\right]e^{-\delta t}dt$$

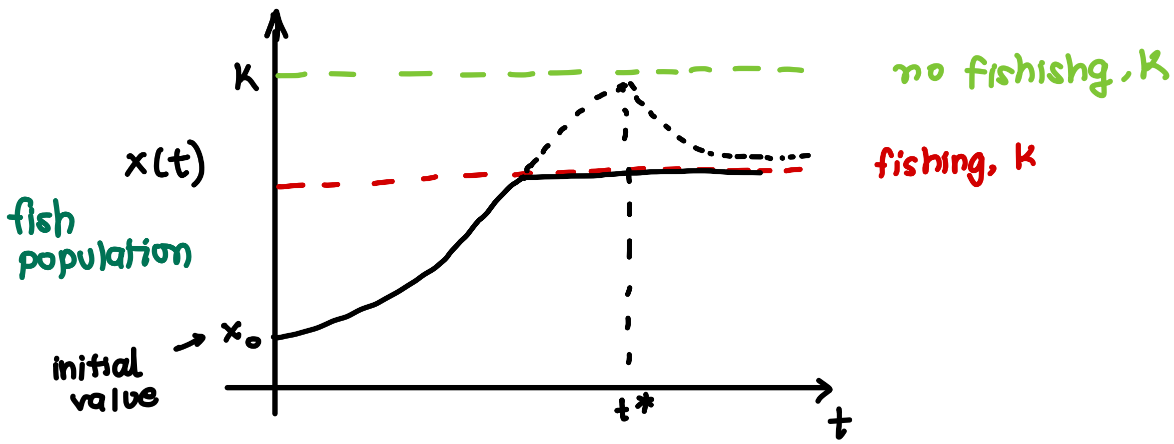
Thus, we can solve (†) and (‡) by first determining the optimal switching time t^* which allows for solving (‡) analytically and plugging the solution $x(t)$ into (†), which then eliminates the constraint and allows us to compute the optimal number of boats.

Step 1. We maximize over the switching time t^*.

Clearly the optimal switching time will depend on the initial value x_0:
If x_0 is larger than the maximum capacity under fishing then it pays off to resume fishing from the very beginning

↑ initial fish population

If however the initial fish population is below the capacity, then one should wait and resume fishing once the fish population has reached the fishable capacity. Waiting longer to further increase the population does not pay off, in particular since future profits are discounted.



The solution of the switched logistic equation at the switching point t^* is continuous but not differentiable because the control variable has a jump discontinuity at t^* and jumps from $b(t^*)=0$ to $b(t^*+\epsilon)=b_0$.

Let us assume that $x_0 < x^*$ and recall that from separation of variables we determined that the fish population has the form

$$x(t) = \frac{Kx_0 e^{\gamma t}}{K + x_0(e^{\gamma t}-1)}$$

When $\boxed{b_0=0}$. We can rewrite this as

$$x(t) = \frac{Kx_0}{Ke^{-\gamma t}+x_0(1-e^{-\gamma t})} \quad \frac{\cancel{e^{\gamma t}}}{\cancel{e^{\gamma t}}}$$

$$= \frac{Kx_0}{(K-x_0)e^{-\gamma t}+x_0}$$

$$= \frac{K}{\left(\frac{K}{x_0}-1\right)e^{-\gamma t}+1} \quad \frac{\cancel{x_0}}{\cancel{x_0}}$$

$$= \frac{K}{1+(K/x_0-1)e^{-\gamma t}}, \quad t \in [0,t^*]$$

solution to logistic eqn in the initial period $[0,t^*]$ without fishing, i.e. $b_0=0$

The optimal switching time is then determined by the condition $x_0(t^*) = x^*$

Solving the equation for t^* yields

$$x^* = \frac{K}{1+(K/x_0-1)e^{-\gamma t^*}}$$

$$x^* + x^*\left(\frac{K}{x_0}-1\right)e^{-\gamma t^*} = K$$

$$x^*\left(\frac{K}{x_0}-1\right)e^{-\gamma t^*} = K - x^*$$

$$e^{\gamma t^*} = \frac{x^*\left(\frac{K}{x_0}-1\right)}{K-x^*}$$

$$\gamma t^* = \log\left[\frac{x^*\left(\frac{K}{x_0}-1\right)}{K-x^*}\right]$$

$$\Rightarrow \quad t^* = \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)+\log\left(\frac{x^*}{K-x^*}\right)\right]$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)-\log\left(\frac{K-x^*}{x^*}\right)\right]$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)-\log\left(\frac{K}{x^*}-1\right)\right]$$

From pg 119, we found $x^* = K\left(1-\frac{qb_0}{\gamma}\right)$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)-\log\left(\frac{1}{1-\frac{qb_0}{\gamma}}-1\right)\right]$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)-\log\left(\frac{\cancel{1}-\cancel{1}+\frac{qb_0}{\gamma}}{1-\frac{qb_0}{\gamma}}\right)\right]$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)+\log\left(\frac{1-qb_0}{\frac{qb_0}{\gamma}}\right)\right] \quad \leftarrow \text{using } -\log(x) = \log\left(\frac{1}{x}\right)$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)+\log\left(\gamma\left(\frac{1-qb_0}{qb_0}\right)\right)\right]$$

$$= \frac{1}{\gamma}\left[\log\left(\frac{K}{x_0}-1\right)+\log\left(\gamma\left(\frac{1}{qb_0}-1\right)\right)\right]$$

which determines the optimal switching time $t^* = t^*(b_0)$ as a function of the number of boats (via the capacity K, that is a function of b_0).

Step 2. Next, we eliminate the constraint from J, by noting that

$$x(t)=x^* \quad \forall\, t \geq t^*$$

Hence $J(b)=\int_{t^*}^{\infty} b_0\left(pq\,x^* - c\right)e^{-\delta t}\,dt$

$$= b_0\int_{t^*(b_0)}^{\infty}\left(pqK\left(1-\frac{qb_0}{\gamma}\right)-c\right)e^{-\delta t}\,dt$$

$$= \frac{b_0}{-\delta}\lim_{A\to\infty}\left[\left(pqK\left(1-\frac{qb_0}{\gamma}\right)-c\right)e^{-\delta t}\right]_{t=t^*(b_0)}^{A}$$

$$= -\frac{b_0}{\delta}\left(pqK\left(1-q\frac{b_0}{\gamma}\right)-c\right)\lim_{A\to\infty}\left(\underset{0 \text{ for } \delta>0}{\cancel{e^{-\delta A}}}-e^{-\delta t^*(b_0)}\right)$$

$$= -\frac{b_0}{\delta}\left(pqK\left(1-\frac{qb_0}{\gamma}\right)-c\right)\left(-e^{-\delta t^*(b_0)}\right)$$

$$= \frac{b_0}{\delta}\underbrace{\left(pqK\left(1-\frac{qb_0}{\gamma}\right)-c\right)}_{\text{if } >0 \text{ then } J(b)>0}e^{-\delta t^*(b_0)}$$

The profit function is non-negative when $pqK\left(1-\frac{qb_0}{\gamma}\right)>c$ with c = total cost per boat.

Then rearranging this inequality for b_0 we arrive at

$$1-\frac{qb_0}{\gamma} > \frac{c}{pqK}$$

$$1-\frac{c}{pqK} > \frac{qb_0}{\gamma}$$

$$b_0 < \frac{\gamma}{q}\left(1-\frac{c}{pqK}\right)$$

which implies that for $\boxed{0 \leq b_0 \leq \frac{\gamma}{q}\left(1-\frac{c}{pqK}\right)}$

the function $J(b)$ is bounded from below by 0 and has a unique maximum by Rolle's theorem.

THE END