L3 Astro - Section 6 - Galaxies

Nick Kaiser

November 28, 2020

Contents

1 Types of Galaxies **3**
- 1.1 Hubble's 'Tuning Fork' Diagram . . . 3
- 1.2 Dwarf Galaxies . . . 4
- 1.3 Demographics: the galaxy luminosity function . . . 4

2 Newtonian gravity **4**
- 2.1 Inertial frames and Galilean relativity . . . 4
- 2.2 The gravitational force and acceleration . . . 5
- 2.3 The kinetic energy T . . . 6
- 2.4 The gravitational binding energy . . . 6
  - 2.4.1 The gravitational binding energy is $U = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i$ . . . 6
  - 2.4.2 The gravitational binding energy is $U = -\frac{1}{2}\sum_i \sum_{j \neq i} Gm_i m_j / |\mathbf{r}_i - \mathbf{r}_j|$ . . . 7
  - 2.4.3 The gravitational potential and acceleration fields $\phi(\mathbf{r})$ and $\mathbf{g}(\mathbf{r})$ . . . 8
  - 2.4.4 Potential and gravity for a continuous density field $\rho(\mathbf{r})$ . . . 8
- 2.5 Poisson's equation and Gauss's law . . . 8
- 2.6 The potential energy in terms of the gravity . . . 10
- 2.7 The Newtonian gravitational stress tensor . . . 10

3 Galactic Dynamics **11**
- 3.1 The Virial Theorem . . . 12
- 3.2 Particle discreteness effects in stellar dynamics . . . 13
  - 3.2.1 The relaxation time . . . 13
  - 3.2.2 Dynamical friction . . . 15
  - 3.2.3 Gravo-thermal instability . . . 16
- 3.3 Collisionless stellar dynamics . . . 17
  - 3.3.1 Description of collisionless particles vs. fluids: the phase-space density . . . 17
  - 3.3.2 The fluid limit . . . 18
  - 3.3.3 Equations of motion . . . 19
  - 3.3.4 Properties of the phase-space density . . . 19
  - 3.3.5 The collisionless Boltzmann (or Vlasov) equation and Liouville's theorem . . . 20
  - 3.3.6 Some aspects of Liouville's theorem . . . 22
  - 3.3.7 Coarse grained phase-space density . . . 26
  - 3.3.8 Violent Relaxation . . . 27
  - 3.3.9 Taking moments of the Vlasov equation . . . 27
  - 3.3.10 The zeroth moment: the continuity equation in 3D . . . 28
  - 3.3.11 The first moment: Jeans's equation . . . 29
  - 3.3.12 The Euler equation . . . 30
- 3.4 The collisional Boltzmann equation . . . 31
  - 3.4.1 Statistical mechanical entropy - Boltzmann's H-theorem . . . 33
  - 3.4.2 Application of the H-theorem to gravitating systems . . . 36
  - 3.4.3 Maximum entropy image reconstruction . . . 36

4 Galaxy evolution **37**

List of Figures

1 The Hubble 'tuning fork' diagram . . . 3
2 The galaxy luminosity function . . . 4
3 Gravitational binding energy . . . 7
4 Proof of Poisson's equation . . . 9
5 Gauss's law . . . 9
6 The energy density of the gravitational field . . . 10
7 The 2-body relaxation time . . . 14
8 Dynamical friction . . . 15
9 Penrose on gravitational entropy . . . 17
10 Liouville's theorem . . . 20
11 Conservation of mass or particle number for a fluid . . . 21
12 The mapping of phase-space under time evolution . . . 23
13 Adiabatic invariance . . . 24
14 The 1911 Sovay conference . . . 24
15 Lorentz invariance of phase-space density . . . 25
16 The coarse grained phase space density . . . 27
17 A 2-body collision . . . 32
18 The assumption of 'molecular chaos' . . . 35
19 MAXENT image reconstruction . . . 37

1 Types of Galaxies

1.1 Hubble's 'Tuning Fork' Diagram

- Hubble classified galaxies by their visual characteristics
  - 2 main classes: Elliptical and spiral galaxies
- Ellipticals:
  - surface brightness: featureless - ellipsoidal - surface brightness: $\Sigma \sim \exp\left(-(r/r_*)^{1/4}\right)$
    * that's for a spherical elliptical (E0)
    * we can make an ellipsoidal model as $\Sigma \sim \exp\left(-(r_i M_{ij} r_j)^{1/8}\right)$
    * hard to say from one object if we're looking at a prolate (needle-like) or oblate (burger-like) spheroid but statistics and kinematics (below) suggest they are mostly oblate
  - kinematics:
    * disordered motion – like a collisionless gas with some phase-space distribution
    * some rotation – stronger in lower luminosity ellipticals
  - little or no star formation - 'red and dead' - 'passively' evolved populations
  - tendency to be found in clusters
  - Hubble classification: 0 = round $\Rightarrow$ 9 = highly flattened
- Spirals:
  - highly flattened
    * overall exponential surface brightness $\Sigma \sim \exp(-r/r_*)$
    * with spiral and/or 'barred spiral' patterns superposed
  - disks usually blue *Rrightarrow* star formation – 'active' population
  - tendency to avoid clusters
  - classified by tightness of spiral structure $a, b, c \Leftarrow$ from tightly to loosely wrapped
  - further classified into barred (SBa,b,c) and non-barred (Sa,b,c)
    * hence 'tuning fork' – handle=E-gals; tines=S-gals
- S0: disk-like but featureless 'lenticular' galaxies placed in the transition region
- initially imagined to be an evolutionary sequence – hence the terminology "early type" and "late type" for ellipticals and spirals respectively

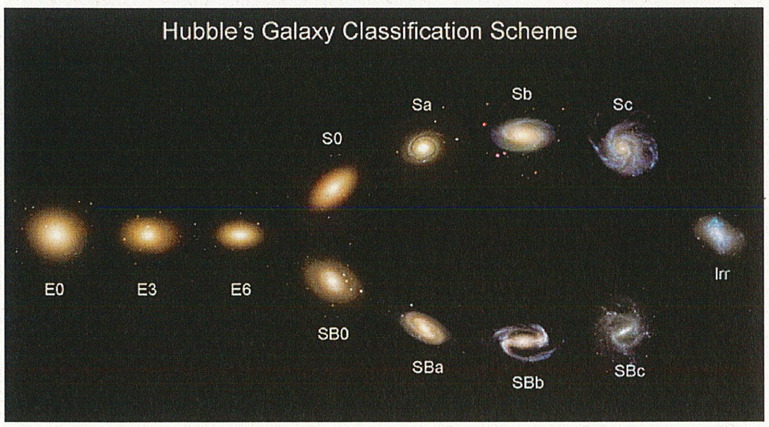

Figure 1: Hubble's classification scheme. Often called the 'tuning fork diagram'.

1.2 Dwarf Galaxies

- A type of galaxy missing from Hubble's diagram is the *dwarf irregular*
  - low luminosity
  - active stellar population
- Later realised that low luminosity *ellipticals* are different from more massive ones
  - different scaling of surface brightness with luminosity (Kormendy)
  - hence a 4th type of galaxy *dwarf ellipticals*

1.3 Demographics: the galaxy luminosity function

- within each morphological classification there is wide range of luminosities
- described by the differential *luminosity distribution function*
- counts ⇒ well fit by a *'Schechter function'*: $\boxed{dn/dL \propto L^{-\alpha} \exp(-L/L_\star)}$
  - number per unit spatial volume per unit luminosity
  - power-law at low luminosities – slope depends on galaxy type
  - exponential 'cut-off' above the 'knee' $L_\star$ ⇐ *characteristic luminosity*
- the *number* of galaxies may be formally infinite – i.e. growing without limit as the observations probe fainter since $dN(> L)/d\log L \propto L^{1-\alpha}$ – but the total luminosity is finite – with most of the luminosity in galaxies with $L \sim L_\star$

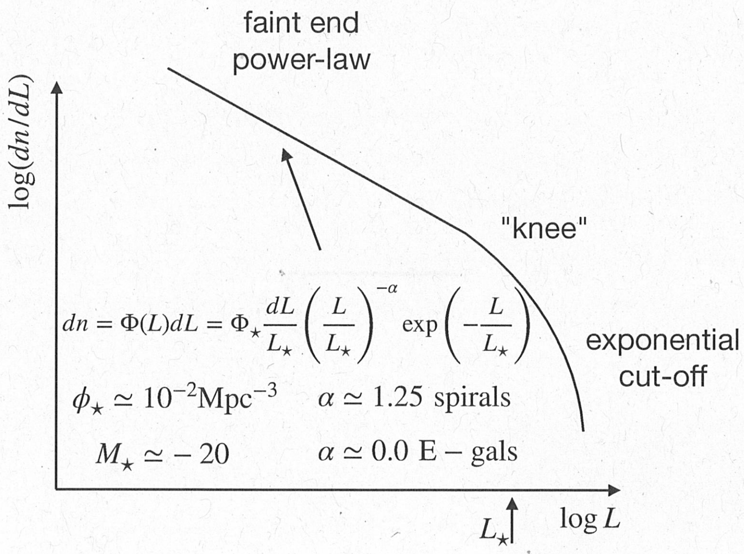

Figure 2: Galaxy (differential) *luminosity function:*. Massive *redshift surveys* (like SDSS) have been performed to determine the 3D distribution of galaxies. The counts of galaxies as a function of luminosity are well described by what's called a *'Schechter function'* after pioneering work by Paul Schechter. It is a power law at low luminosities – with a formal divergence of the number of galaxies – and with a sharp exponential cut-off above what's called the *'characteristic luminosity'* $L_\star$. The characteristic absolute magnitude is $M_\star \simeq -20$, which is about the absolute magnitude of the Milky Way. Elliptical galaxies have a flatter 'faint-end'; with $\alpha \simeq 0.0$.

2 Newtonian gravity

We will now review the essentials of Newtonian gravity.

2.1 Inertial frames and Galilean relativity

- Newtonian dynamics in general (and Newtonian gravity in particular) plays out in a space-time arena where there is an *absolute time* t measured by universal clocks and positions $\mathbf{r}$ measured by rulers
- there is, however, no absolute system of spatial coordinates as there is for time

- instead Newton's laws obey the *Galilean principle of relativity:*
  - the laws of physics are the same in any one of a set of *inertial frames* that are in *uniform linear motion* with respect to one another
  - there is no one inertial frame that is special, and with respect to which one can determine one's absolute state of motion
  - Newton's laws embody this principle – if they are obeyed in one inertial frame then they are obeyed in any inertial frame
- differently *accelerating* frames, however, are distinguishable
- neither does the principle extend to relative *uniform rotation*
  - there is an absolute non-rotating frame
  - we can sense whether or not we are non-rotating from stresses in our bodies, motion of test particles (Coriolis effect)
  - this coincides with the frame of reference in which distant stars appear not to rotate on the sky – something that goes by the name of 'Mach's principle'

2.2 The gravitational force and acceleration

- according to Newton, the gravitational force is an inverse square attraction that acts instantaneously between all massive particles
  - he thought this preposterous, but refused to postulate the mechanism by which this influence was transmitted – *"hypotheses non fingo"*
- for a collection of point-like particles of mass m_i and positions $\mathbf{r}_i$
  - where i is an index that labels the particles

  the force on the i^{th} particle is
  - $$\boxed{\mathbf{F}_i = G m_i \sum_{j \neq i} m_j \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3}}$$
- which is *linear* in the sense that the force is the linear sum of the forces from the other particles
- and the acceleration $\mathbf{a}_i = \mathbf{F}_i / m_i$ is independent of m_i and, according to Galileo, is independent of the composition of the accelerated object
  - all objects accelerate the same way under gravity

  this is called the *Galilean principle of equivalence.*
- Newtonian dynamics is superficially similar to low-velocity electrodynamics in that motion of charged particles obey an inverse square law
  - $\mathbf{F}_{12} = G m_1 m_2 \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \quad \leftrightarrow \quad \mathbf{F}_{12} = \epsilon_0^{-1} q_1 q_2 \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3}$
  - in both cases, one can think of particle 2 producing a *force field*
    - $\mathbf{g}$ in the case of gravity
    - $\mathbf{E}$ in the case of electrodynamics
  - that acts on particle 1
  - in general one could construct an inverse square law of gravity with 3 distinct masses
    - an *active* gravitational mass that says how much $\mathbf{g}$ the 2nd particle creates
    - a *passive* gravitational mass that says how much force a given $\mathbf{g}$ generates, and
    - the *inertial mass* which, in $\mathbf{F} = m\mathbf{a}$, tells us how much acceleration a given force produces

  - if one wants Newton's third law (equal and opposite reaction) to hold then the active and passive masses must be the same up to a constant
  - but the passive mass could differ from the inertial mass, as is the case in electrodynamics
  - but Galileo's experiment and later versions thereof show that, for gravity, these also are equal, up to a constant
    * so gravity is like electrodynamics, but where there is, effectively, a *universal charge-to-mass ratio* for all particles
  - and it is this this fact that makes it possible to believe that gravity is not really a force at all but a manifestation of curvature of space as in Einstein's general relativity

- both gravity and electrodynamics obey the Galilean principle of relativity but only gravity obeys the Galilean principle of equivalence

- another feature of Newtonian gravity is that, while it involves a field $\mathbf{g}$ (or the gravitational potential field φ) it is *acausal* (unlike Maxwell's electromagnetism) in that changes in the configuration of particles is communicated *instantaneously* to the other particles

2.3 The kinetic energy T

- The energy of a gravitating system has two components: the *kinetic energy* (KE) and the *potential* or *binding* energy (PE)

- of these, the kinetic energy is straightforward:
  - it is the sum over particles of their individual kinetic energies
  - it is often denoted by T:
  - $$\boxed{T = \sum_i m_i |\mathbf{v}_i|^2/2}$$

- the KE thus defined depends on the inertial frame
  - however, it is the sum two terms: one giving the energy of the particles relative to the centre of mass frame
  - which *is* frame independent
  - and the other being the KE of the system as a whole

2.4 The gravitational binding energy

- the potential energy is more complex as there are various ways to express it
  - one is as the sum of forces dotted with positions
  - another is as a pairwise sum of shared potential energies (like particles connected by springs)
  - which can be expressed as an integral of the density times a gravitational potential field
  - and which can also be expressed, if one likes, purely as an integral involving the gravity alone

2.4.1 The gravitational binding energy is $U = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i$

- one very useful expression for the gravitational binding energy
  - it is what appears in the virial theorem (see below)

  and which we will denote by U, is

  - $$\boxed{U = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i}$$

  the proof of this is given in the caption of figure 3

- it follows from asking how much energy would be released if we were to assemble a collection of particles into a static configuration by bringing them in from infinity
- the total energy for a non-static configuration is then obtained simply by adding the kinetic energy

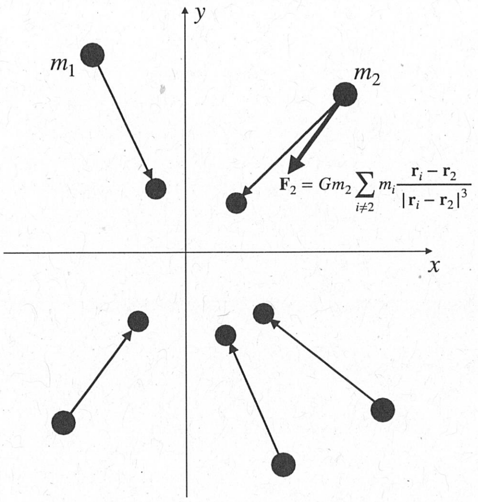

Figure 3: One expression for the gravitational binding energy U of a collection of particles is as the sum over particles of their position $\mathbf{r}$ – relative to the origin of spatial coordinates – dotted with the gravitational force from all the other particles: $U = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i$. To prove this we consider a succession of different configurations where the particles have positions equal to some scale factor $a \geq 1$ times their final positions and we ask how much energy is released as the configuration contracts. This is $dW = \sum_i \mathbf{F}_i(a) \cdot d\mathbf{r}_i(a)$. But $\mathbf{F}_i(a) = \mathbf{F}_i(1)/a^2$ and $\mathbf{r}_i(a) = a\mathbf{r}_i(1)$ and hence $d\mathbf{r}_i = \mathbf{r}_i(1)da$ so $dW = (\sum_i \mathbf{r}_i(1) \cdot \mathbf{F}_i(1)) \times da/a^2$. If we start at $a = \infty$, the energy released is $W = \int_\infty^1 dW = (\sum_i \mathbf{r}_i(1) \cdot \mathbf{F}_i(1)) \times [-1/a]_\infty^1 = -\sum_i \mathbf{r}_i(1) \cdot \mathbf{F}_i(1)$. Conservation of energy requires – as we are considering initial and final configurations with no kinetic energy – that $U = -W = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i$.

2.4.2 The gravitational binding energy is $U = -\frac{1}{2} \sum_i \sum_{j \neq i} Gm_i m_j / |\mathbf{r}_i - \mathbf{r}_j|$

- A more common expression of the binding energy is as a pairwise sum of the potential for a pair of particles $-Gm_1m_2/|\mathbf{r}_1 - \mathbf{r}_2|$
  - but with a factor 2 to account for the fact that the energy is 'shared' between the particles in the pair
- we can justify this starting from $U = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i$ as follows:
  - switching labels on the particles $i \leftrightarrow j$ has no effect, so
  - $U = \sum_i \sum_{j \neq i} Gm_i m_j \mathbf{r}_i \cdot (\mathbf{r}_j - \mathbf{r}_i)/|\mathbf{r}_j - \mathbf{r}_i|^3 = \sum_j \sum_{i \neq j} Gm_i m_j \mathbf{r}_j \cdot (\mathbf{r}_i - \mathbf{r}_j)/|\mathbf{r}_j - \mathbf{r}_i|^3$
  - but $\sum_j \sum_{i \neq j} = \sum_i \sum_{j \neq i}$, since both are just summing over the non-diagonal squares on the 'chess-board', and using $(\mathbf{r}_i - \mathbf{r}_j) = -(\mathbf{r}_j - \mathbf{r}_i)$ the latter expression is
  - $U = -\sum_i \sum_{j \neq i} Gm_i m_j \mathbf{r}_j \cdot (\mathbf{r}_j - \mathbf{r}_i)/|\mathbf{r}_j - \mathbf{r}_i|^3$
  - and averaging this with the original expression gives
  - $U = \frac{1}{2} \sum_i \sum_{j \neq i} Gm_i m_j \frac{(\mathbf{r}_i - \mathbf{r}_j) \cdot (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_j - \mathbf{r}_i|^3}$ or
- $$\boxed{U = -\frac{1}{2} \sum_i \sum_{j \neq i} \frac{Gm_i m_j}{|\mathbf{r}_j - \mathbf{r}_i|}}$$
  - This is the same as the potential energy for a set of particles connected by springs with potential energy – as a function of length r – $Gm_i m_j / r$
  - the factor $1/2$ coming in as the energy in each spring is 'shared' between the two particles that it connects
  - or, equivalently, because the double sum $\sum_i \sum_{j \neq i}$ counts each pair of particles, and therefore each spring, twice
- this form for the binding energy makes transparent the fact that the gravitational binding energy is always negative

2.4.3 The gravitational potential and acceleration fields $\phi(\mathbf{r})$ and $\mathbf{g}(\mathbf{r})$

- we can also express the binding energy as
  - $$\boxed{U = \frac{1}{2}\sum_i m_i \phi(\mathbf{r}_i)}$$

  where the *gravitational potential* is defined to be
  - $$\boxed{\phi(\mathbf{r}) = -\sum_j G m_j / |\mathbf{r}_j - \mathbf{r}|}$$

  and whose gradient is minus the *gravitational acceleration* (or just the *gravity*) vector
  - $$\boxed{\mathbf{g}(\mathbf{r}) = -\boldsymbol{\nabla}\phi(\mathbf{r})}$$

2.4.4 Potential and gravity for a continuous density field $\rho(\mathbf{r})$

- We can make the transition to a continuous mass distribution by writing the *mass density* for a set of particles as $\rho(\mathbf{r}) = \sum_i m_i \delta^{(3)}(\mathbf{r} - \mathbf{r}_i)$
  - with $\delta^{(3)}(\mathbf{r})$ the 3-dimensional *Dirac delta-function*
    * which can be thought of as e.g. the the limit of a small normalised Gaussian:
    * $\delta^{(3)}(\mathbf{r}) = \lim_{\sigma \to 0} (2\pi\sigma^2)^{-3/2} \exp(-|\mathbf{r}|^2/2\sigma^2)$
    * though there are many other possibilities, of which perhaps the simplest is the limit of a small normalised 'box-car'
      · whose 1D version is: $\delta(r) = \begin{cases} \epsilon^{-1} & \text{if } |r| < \epsilon/2 \\ 0 & \text{otherwise} \end{cases}$
      · and whose 3D version is just the product $\delta^{(3)}(\mathbf{r}) = \delta(x)\delta(y)\delta(z)$
  - and which has the fundamental property that, for any function (or field) $f(\mathbf{r})$
    * $$\boxed{\int d^3r' f(\mathbf{r}')\delta^{(3)}(\mathbf{r}' - \mathbf{r}) = f(\mathbf{r})}$$

- and then considering a continuous mass distribution to be the same as a very finely distributed set of point masses in the limit that $m \to 0$

- so we have
  - $$\boxed{U = \frac{1}{2}\int d^3r \rho(\mathbf{r})\phi(\mathbf{r})}$$
  - with
  - $$\boxed{\phi(\mathbf{r}) = -\int d^3r' G\rho(\mathbf{r}')/|\mathbf{r}' - \mathbf{r}|}$$

- and $\mathbf{g}(\mathbf{r}) = -\boldsymbol{\nabla}\phi(\mathbf{r})$ as before.

2.5 Poisson's equation and Gauss's law

- the expression above for $\phi(\mathbf{r})$ is a solution of *Poisson's equation:*
  - $$\boxed{\nabla^2\phi = 4\pi G\rho}$$

  with *boundary conditions* $\phi \to 0$ as $r \to \infty$

- the validity of Poisson's equation can be established as follows
  - first, consider a single point mass m at $\mathbf{r} = 0$
  - the potential is $\phi = -Gm/|\mathbf{r}|$

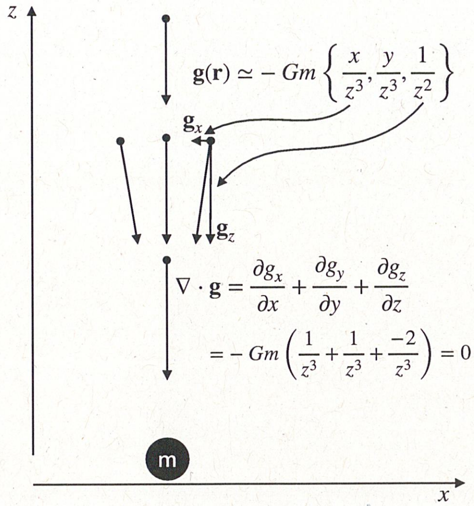

Figure 4: Proof that $\nabla \cdot \mathbf{g} = 0$ outside a point mass. Consider a point $\mathbf{r}_0 = \{0, 0, z\}$ along the z-axis above a point mass m. In the vicinity of that point, the gravity is, to first order in the transverse displacements x and y given by $\mathbf{g}(\mathbf{r}) \simeq -GM\{x/z^3, y/z^3, 1/z^2\}$, whose divergence vanishes. A vector field $\mathbf{f} \propto \hat{\mathbf{r}}/|\mathbf{r}|^2$ like $\mathbf{g}$ is the simplest, and perhaps archetypical, example of a *divergence free flow:* If we think of a fluid with flux density (product of density $\rho(\mathbf{r})$ and velocity $\mathbf{v}(\mathbf{r})$): $\mathbf{f}(\mathbf{r})$ then if we consider concentric spheres, the rate at which fluid is crossing the surface is proportional to the area ($\propto |\mathbf{r}|^2$) times the flux density $\mathbf{f}$ ($\propto 1/|\mathbf{r}|^2$) and is the same for all spheres. Provided we keep supplying fluid at $\mathbf{r} = 0$ (perhaps think of a garden sprinkler, in which case $\rho \propto 1/r^2$ and $\mathbf{v} =$ constant) then there will be no build-up or decrease of density of fluid. This is expressed in the *continuity equation* $\partial\rho/\partial t = -\nabla \cdot \mathbf{f}$.

  - and the gravitational acceleration is $\mathbf{g} = -Gm\mathbf{r}/|\mathbf{r}|^3$
  - from which it follows that $\nabla \cdot \mathbf{g} = 0$ for $\mathbf{r} \neq 0$ (see figure 4)

- but owing to linearity of the force law that means that anywhere that $\rho = 0$ the divergence of the gravity must vanish: the matter elsewhere having no effect
- and hence $\nabla^2\phi = \nabla \cdot \nabla\phi = -\nabla \cdot \mathbf{g} = 0$ also
- so $\nabla^2\phi$ must be proportional to the density, and the constant of proportionality ($4\pi G$) is readily established considering a small uniform density sphere
- applying the *divergence theorem* gives *Gauss's law:* (see figure 5)
  - $$\boxed{\int d\mathbf{A} \cdot \mathbf{g} = -4\pi G \int d^3r\rho}$$
  - so the integral of the outward component of the gravity over the boundary of a volume is $4\pi G$ times the mass enclosed.
- solving Poisson's equation for some given mass density field gives the potential, and hence the gravity
  - though this is arguably of somewhat limited utility given that we can simply write down the potential – and be confident that the result has the proper boundary conditions at infinity
  - it is very useful for establishing useful relations between other quantities as we shall now illustrate

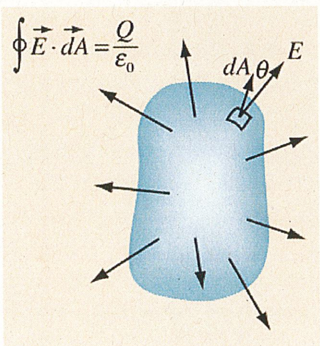

Figure 5: This illustrates (the integral form of) Gauss's law in electromagnetism. It is equivalent the the Maxwell equation $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ by virtue of the divergence theorem. The gravitational version follows from Poisson's equation in precisely the same manner.

2.6 The potential energy in terms of the gravity

- just as in electro-statics, where one can consider the energy to be the sum over charges of the potential
  - i.e. considering the energy to be associated, and localised, with the charges
- or as an integral $\frac{\epsilon_0}{2} \int d^3r |\mathbf{E}|^2$
  - i.e. considering the energy to be associated and localised with the field
- we can express the binding energy entirely in terms of the gravity
  - if we perform the integral $I = \int d^3r |\nabla\phi|^2$ by parts we get, with sensible boundary conditions, $I = -\int d^3r \phi \nabla^2 \phi$
  - which with Poisson's equation is $I = -4\pi G \int d^3r \rho\phi = -8\pi GU$
- $$\boxed{U = -(8\pi G)^{-1} \int d^3r |\mathbf{g}|^2}$$
- so the energy density associated with the gravitational field is $\epsilon = |\mathbf{g}|^2/8\pi G$
- but, as with electrostatics, we cannot 'double-count':
  - either we consider the energy to be associated with the masses
  - or with the field
  - but not both

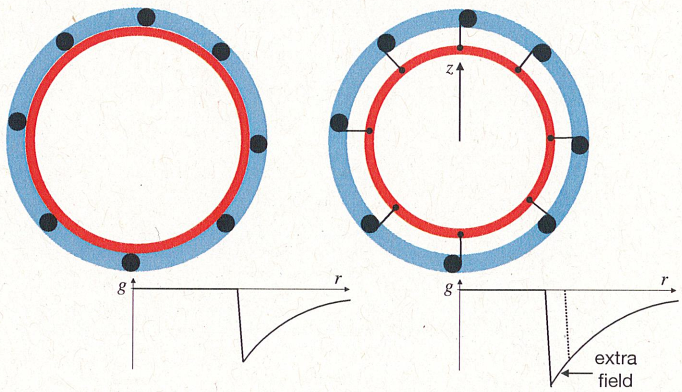

Figure 6: Inside a light rigid shell (blue) is suspended (against its own self-gravity) a massive shell (red). This is lowered by winches which gain energy. This creates new **g**-field in the region that was previously inside the shell (and therefore field-free). The energy gained is $|\mathbf{g}|^2/8\pi G$ times the volume it occupies: hence the energy density of the newly created field is $\epsilon = -|\mathbf{g}|^2/8\pi G$. The cable at the top is carrying +ve z-momentum downwards: a negative flux of z-momentum. For momentum to be conserved, there must be a positive flux of z-momentum in the **g**-field.

2.7 The Newtonian gravitational stress tensor

This section is not absolutely essential for what follows, but it is highly useful to realise that, just as with electromagnetism, there is a *momentum flux density* in the gravitational field. It is directly analogous to the *'Maxwell stress'* tensor in EM. It has properties that might appear strange; there is an immense flux of momentum upwards out of the earth in the room where you are sitting. But in empty space it is divergence-free; so there is no build-up of momentum anywhere, except in massive bodies where there is a divergence, and this is how the gravitational field transfers momentum between massive bodies.

- In electromagnetism, we have the Maxwell stress tensor
  - $\mathbf{T} = \epsilon_0(\mathbf{EE} - \mathbf{I}|\mathbf{E}|^2) + \mu_0^{-1}(\mathbf{BB} - \mathbf{I}|\mathbf{B}|^2)$

  or, in component form,
  - $T_{ij} = \epsilon_0(E_iE_j - \frac{1}{2}E^2\delta_{ij}) + \mu_0^{-1}(B_iB_j - \frac{1}{2}B^2\delta_{ij})$

and which is the *momentum flux density* of the field.

- momentum being a vector quantity, its flux density is necessarily a tensor
- this describes the transport of momentum by EM waves
- and also the flux of momentum in the field between two capacitor plates that are being prevented from coming together by springs: it provides the continuity of momentum needed as the plates are neither gaining nor losing momentum

- in Newtonian gravity there is a precisely analogous flux of momentum that one can associate with the *gravitational* field. It is given by
  - $T_{ij} = -(8\pi G)^{-1}(g_i g_j - \frac{1}{2}g^2\delta_{ij})$
- Maxwell was the first to point out that, at the surface of the earth this is positive
  - so if we sit at the N-pole (i.e. +ve z) there is a flux of z-momentum upwards
  - the caption to figure 6 describes how this flux balances the negative momentum flux density in the cables supporting the mass shell
- and it's value is 32,000 tons per square inch!
  - this deterred Maxwell from developing a Lorentz invariant gauge-field theory for gravity
  - he didn't believe that the underlying mechanism of space – what he called the 'underworld' – could be strong enough to withstand such a stress
- Some problems to consider:
  - Q: Show that the 3 components of the divergence of the flux density: $\partial T_{ij}/\partial x_j$ (outside the Sun for instance) vanishes in empty space
    * but that it is non-zero inside the planets in the solar system
    * and that it properly accounts for the changing mechanical momentum of those objects
  - Q: The strength of gravity falls off as $1/r^2$ so the components of T_{ij} fall off as $1/r^4$
    * thus it would seem that, in empty space, the flux of momentum across spheres around a point mass (area times flux density) is decreasing
    * is this a problem? Does this violate conservation of momentum? Where is all of that momentum going?

3 Galactic Dynamics

- The subject of *galactic dynamics* is concerned with the modelling of galaxies (and other systems such as galaxy- and star-clusters) as self-gravitating systems composed of very large numbers of stars.
- The velocities of stars galaxies are typically on the order of $\sim 100 - 300$km/s
- Velocities of galaxies in galaxy clusters are typically $\sim$ 1000km/s, but still much less than $c = 300,000$km/s
- so we are well justified in using *Newtonian gravity* in this work

The 'bible' is the graduate level textbook by James Binney and Scott Tremaine. However, the introductory parts of the chapters relevant for us are mostly quite accessible.

3.1 The Virial Theorem

An often used method for estimating the mass of a self-gravitating collection of stars (in a galaxy or a star cluster) or galaxies (in a cluster of galaxies) is the *virial theorem*. This states that for a bound system that is neither expanding nor contracting the potential energy is (minus) twice the kinetic energy.

- Consider a collection of N particles interacting via their mutual gravitational attraction.
- Let them have positions $\mathbf{r}_i$, where i is an index that labels the particles, and velocities $\dot{\mathbf{r}}_i$.
- Consider the quantity $\boxed{I \equiv \sum_i m_i \mathbf{r}_i \cdot \mathbf{r}_i / 2}$
  - I is (half) the moment of inertia
  - its time derivative is $\dot{I} = \sum_i m_i \mathbf{r}_i \cdot \dot{\mathbf{r}}_i$
  - and its second derivative is $\ddot{I} = \sum_i m_i (\mathbf{r}_i \cdot \ddot{\mathbf{r}}_i + \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i)$
  - so we have
    * $\ddot{I} = U + 2T$
    * where
    * $T \equiv \sum_i m_i v_i^2 / 2$ is the *total kinetic energy* and
    * $U \equiv \sum_i m_i \mathbf{r}_i \cdot \ddot{\mathbf{r}}_i$ is the *gravitational binding energy*
- Now if the system of particles is in a stable state
  - so it is neither expanding nor contracting

  then $\dot{I} = \sum m\mathbf{r} \cdot \dot{\mathbf{r}}$ must vanish
  - since I is a measure of the size of the system
  - though there is actually a hidden assumption here: that we are working in that inertial frame in which there is no motion of the centre of mass
    * were this not the case, one would not expect I to be constant

  and that implies that $\ddot{I}$ must vanish also
  - since, were this not the case, then I would not remain constant

  and this gives the famous *virial theorem* for a stable, gravitationally bound system:
  - $\boxed{2T = -U}$
- Some features of the virial theorem:
  - Since the potential energy varies as the square of the mass while the kinetic energy varies as the first power, if we were to observe velocities to get T/m (assuming for simplicity that all of the masses are the same) and positions to get U/m^2, we can combine these and solve for the particle mass m.
    * A practical handicap of the virial theorem is that it involves 3-dimensional positions and velocities whereas what we usually have access to is 2-dimensional positions and line-of sight velocity. So to make progress it is necessary to assume e.g. spherical sysmmetry, and to assume that the mean squared l.o.s. velocity is 1/3 of the 3D velocity squared. Another possibility is to relate the so-called *'pair-weighted harmonic mean'* – the average of $1/r$ – appearing in U to the equivalent average of projected separations assuming that pair separations are, on average, isotropically distributed.
    * Another handicap, not shared by Jeans equation (see below), is that the model here is that all the mass is in the objects we can observe, and that the masses are the same, or that the relative masses are known. These are somewhat dubious assumptions in general.

- Another key feature of the virial theorem is that $E = T + U$, so
  * $\boxed{E = -T/2}$

  which, since T is necessarily positive, implies that the total energy of a self-gravitating system in virial equilibrium is *negative*
  * that means that gravitating systems have *negative specific heats*:
  * since adding (removing) energy to (from) such a system makes it colder (hotter).
- the virial theorem is similar in physical content to the equation of hydrostatic equilibrium
  * $\nabla P = -g\rho$
  * since the latter implies that $m \times P/\rho$ – which is the mean kinetic energy per particle – is on the order of $m \times gR \sim GMm/R \simeq -U$

3.2 Particle discreteness effects in stellar dynamics

- In a previous section we described how Jan Oort 'weighed' the disk of the MW.
- in doing so he assumed that stars in the disk move under the influence of the gravitational force due to the 'smoothed out' distribution of stars (and dark matter).
  - i.e. he ignored the effect of interactions *collisions* between the individual stars.
  - or, equivalently, he ignored the *'graininess'* of the gravitational potential associated with the discreteness of stars
- in this section we will review three effects that arise from the discreteness of the masses in real gravitating systems
- the first is *dynamical relaxation:* which is the fact that particles will gradually 'lose memory' of their initial velocities
  - an important outcome of which is the *relaxation time*
  - this tells us over what time-scales we can neglect the graininess of matter
  - we will see that in many, but not all, cases it is extremely long, so neglecting graininess can be a neglected
- We then discuss *'dynamical friction:'* the process by which the more massive objects in a gravitating system will tend to sink to the centre
  - this can lead to *mass segregation*
  - though we will see that this generally happens on a time-scale similar to the relaxation timescale
- Finally we will discuss how the fact – revealed by the virial theorem – that gravitating systems have *negative specific heats* leads them to be *thermally unstable*

3.2.1 The relaxation time

- Consider a star passing through a uniform 'dust' of other stars (i.e. a Poisson distribution in space) (see figure 7).
- Its path will not be exactly straight as it will suffer gravitational interactions and get deflected
- Q: how long will it take for the star to 'forget' about its initial velocity? This is the *relaxation time*
  - consider first a single interaction with another star (mass m) that it passes with impact parameter $\sim b$
    * time-scale for interation is $\boxed{t \sim b/v}$
    * change in velocity $\Delta v_1 =$ acceleration $\times$ time $\sim (Gm/b^2) \times (b/v)$
    * so the deflection angle is $\Delta\theta_1 \sim \Delta v_1/v \sim Gm/bv^2$

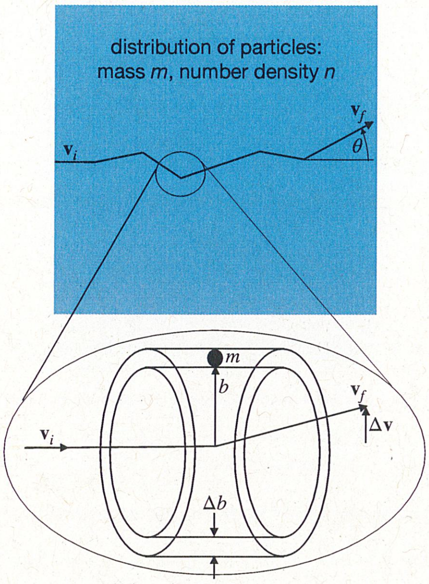

Figure 7: Illustration of how we calculate the relaxation time. The idea is that a particle fired initially in the x-direction say will suffer random interactions with particles it passes. Each interaction will cause a – assumed small – deflection of the direction of motion. The trajectory will then be a random walk, where the mean deflection angle – a 2-vector living in the $y-z$ plane – will be zero, but it will have some non-zero mean squared value $\langle\theta^2\rangle$ which will grow linearly with time. The argument presented in the text is to first compute $\langle\theta^2\rangle$ resulting from those interactions where the impact parameter lies in some logarithmic integral around b. A single interaction gives $\langle\Delta\theta_1^2\rangle \propto 1/b^2$; the close encounters having the most effect. But the number of such interactions, in travelling a certain distance – the size of the system, for instance – is proportional to b^2 so the contribution to $\langle\theta^2\rangle$ is independent of b. In fact, it is $\langle\theta^2\rangle_b \sim 1/N_\star$ where $N_\star$ is the number of stars in the system. To get the full result we need to sum over the different log-intervals of b, with a cut-off at a minimum value such that such interactions individually produce large deflection.

    * but for a typical particle $v^2 \simeq GM/R$ – where M is the total mass and R is the size of the system – so the angular deflection for a single collision is
    * $\boxed{\Delta\theta_1 \sim m/M \times R/b}$
  - now consider the cumulative effect of many deflections: these will deflect the star in random directions from its path, so the deflections will perform a random walk. We need to sum the squares of the individual deflections to get the cumulative squared deflection
  - in 1 orbit crossing the number of encounters with impact parameter $\sim b$ is $N_{1-\mathrm{orb}} \sim Rb^2 n \sim Rb^2\rho/m$
    * where $\rho \sim nm$ is the mass density

    and the mean square deflection is $(\Delta\theta)^2_{1-\mathrm{orb}} \simeq N_{1-\mathrm{orb}}(\Delta\theta_1)^2 \sim (Rb^2\rho/m)(mR/Mb)^2 = m/M$
  - or, letting $N_\star = M/m$ denote the number of stars in the system
  - $\boxed{(\Delta\theta)^2_{1-\mathrm{orb}} = 1/N_\star}$
    * which is small in any system composed of many stars
    * and as a consequence it takes $\sim N_\star$ orbits to obtain a net root-mean-squared deflection of order 1-radian.
  - that was just the effect of one 'logarithmic interval' – say one e-folding – of impact parameter b
  - the full deflection including all impact parameters is larger by a factor
  - $\boxed{\Lambda \equiv \log(b_{\mathrm{max}}/b_{\mathrm{min}})}$
    * where $b_{\mathrm{max}} \sim R$, the size of the system
    * and $b_{\mathrm{min}} \sim$ the impact parameter that gives $\Delta\theta_1 \sim 1$
  - this reduces the relaxation time, but only by a logarithmic factor

- thus we obtain – to order of magnitude – the *relaxation time:*
  - $\boxed{t_{\mathrm{rel}} \sim N t_{\mathrm{dyn}}/\Lambda}$
  - which implies
  - $\boxed{t_{\mathrm{rel}} \gg t_{\mathrm{dyn}} \quad \text{if} \quad N \gg 1}$

- The conclusion is that for any self-gravitating system containing a large number of particles one can ignore, over a dynamical time-scale at least, the effect of the 'graininess' of the matter distribution.

- For galaxies, containing of order billions of stars, the relaxation time is enormously longer than the age of the universe.

- For globular clusters, however, the relaxation time can be of order the age of the universe so we might expect to see some effects of this.

3.2.2 Dynamical friction

A closely related phenomenon is that of *dynamical friction* (see figure 8) where we consider the drag on a massive object passing through a 'sea' of other objects.

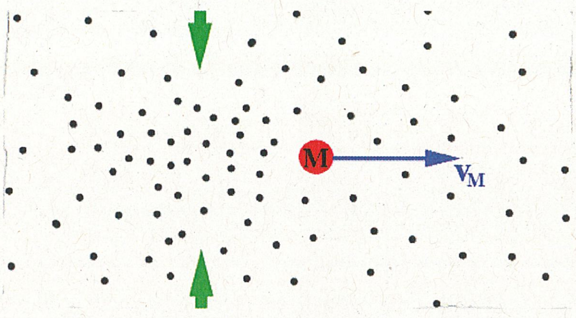

Figure 8: A massive particle moving through a uniform density background will attract particles and induce a 'wake' behind it. This overdensity will exert a force on the particle which will slow it down. One can estimate the time-scale for this process in a manner very similar to that for the relaxation time. The result is that a relatively massive particle containing a fraction f of the total mass will suffer a drag: $\dot{\mathbf{v}} = -\Gamma \mathbf{v}$ with decay rate $\sim \Lambda f / t_{\rm dyn}$. Figure courtesy of Frank van den Bosch.

- let the mass of the particle be M and its velocity be v
- as before, consider the matter within some distance b
  - so the time-scale is $\delta t \sim b/v$
- this material will feel a gravitation acceleration $g \sim GM/b^2$
- and so will gain velocity (an 'impulse') $\delta v \sim g \delta t \sim GM/bv$
- and therefore move a distance $\delta r \sim \delta v \delta t \sim GM/v^2$ towards the path of the particle
  - which we note is independent of b
- this will cause a fractional density enhancement $\delta \rho / \rho \sim \delta r / b \sim GM/bv^2$
- the retarding acceleration (or 'drag force') is therefore
  - $\dot{v} \sim -G\delta m/b^2 \sim -G(b^3 \delta \rho)/b^2 \sim -Gb\rho \times \delta\rho/\rho \sim -G^2 \rho M/v^2$
- or, writing $\dot{v} = -\Gamma v$, the *velocity decay rate* is
  - $\Gamma \sim G^2 \rho M / v^3$
- which we note is also independent of b
  - as with the relaxation time calculation, including the effect of all impact parameters serves only to increase the decay rate by a logarithmic factor:
  - $\boxed{\Gamma \sim \Lambda G^2 \rho M / v^3}$
- If the particle belongs to a sub-population that contains a fraction f of the total mass of the system, i.e. $M \sim f \rho R^3$, the decay rate is $\Gamma \sim f \Lambda G^2 \rho^2 R^3 / v^3$
  - or, using $G\rho = 1/t_{\rm dyn}^2$, the decay rate is
  - $\boxed{\Gamma \sim \Lambda f / t_{\rm dyn}}$
  - i.e. to order of magnitude the same as the inverse of the relaxation time for a system with $N \sim 1/f$ particles
- The decay rate is proportional to the mass of the particle.
  - this means that in a globular cluster the more massive particles will have higher orbital decay rates and will sink to the centre
  - this will lead to segregation of the stars by mass
- dynamical friction has many other applications

- galaxies contain, at their centres, massive black holes – but galaxies merge with each other. It is dynamical friction that causes the black holes to spiral in to the centres of the resulting galaxy
- it can important for galaxies in clusters of galaxies and is believed to be what results in the formation of giant, bloated massive central galaxies often found in such clusters

3.2.3 Gravo-thermal instability

Another type of relaxation can occur in globular clusters.

- Many stars exist in binary systems.
- In a 'hard' binary – one in which the orbital velocity is higher than the velocity dispersion of stars orbiting in the cluster – the effective temperature $T \sim mv^2/k$ will be higher for the stars in the binary than for the general population of stars
- it seems reasonable to assume that energy will tend to flow from hotter to cooler systems
- if so, the hard binaries will, in the process of interacting with other stars passing by, lose energy
- but as the total energy of a binary – kinetic plus potential – is negative this means that the binaries will become even more tightly bound

For any self-gravitating – or 'virialised' system (i.e. obeying the virial theorem, so $2T + U = 0$ and therefore $E = T + U < 0$) – the effective *specific heat* – i.e. how much does the energy change with temperature – is *negative*.

And self-gravitating systems are, by virtue of these negative specific heats inherently unstable.

This means that there is no sensible long-term stable equilibrium for a gravitating system.

This was based on the *assertion* that heat flows from 'hot' to 'cold' systems. But this can be justified by calculating the change in *entropy*. As we show later, it increases for such a transfer of energy.

This has the interesting implication that, thermodynamically, for a system with a fixed number of stars and total energy, it is entropically more favoured to have a situation where there is one very tight binary containing essentially all of the gravitational binding energy surrounded by a diffuse lightly bound halo comprised of all the other stars than to have a well mixed system (see Frank Shu's introductory text for a nice discussion of this).

This is the opposite of what happens with e.g. a collisional gas where the highest entropy is for a spatially uniform and 'well-mixed' state.

- for gravitating systems the entropy is higher the more *inhomogeneous* the mass distribution
- this tells us something interesting about the initial state of the universe (as emphasised by Roger Penrose - see figure 9)
  - the cosmic microwave background tells us that, in the early phases of the hot big bang, the plasma was in thermal equilibrium
- so, microphysically, the *matter* had close to the maximum possible entropy it could have
  - which may seem problematic, since the 2nd law of thermodynamics tells us entropy is supposed to increase with time

  but it also tells us that the mass distribution was very close to being homogeneous

  - which, *gravitationally,* is a state of very *low* entropy

  so the initial state was, in fact, a state of exceptionally low total entropy
- and very different from what we think the final state will be
  - it turns out that the entropy of a single super-massive black hole (the 'Hawking-Bekenstein' entropy) is greater than the entire entropy of the CMB

It should be noted that the system may not be able to reach that entropically most favoured state as the systems may effectively decouple.

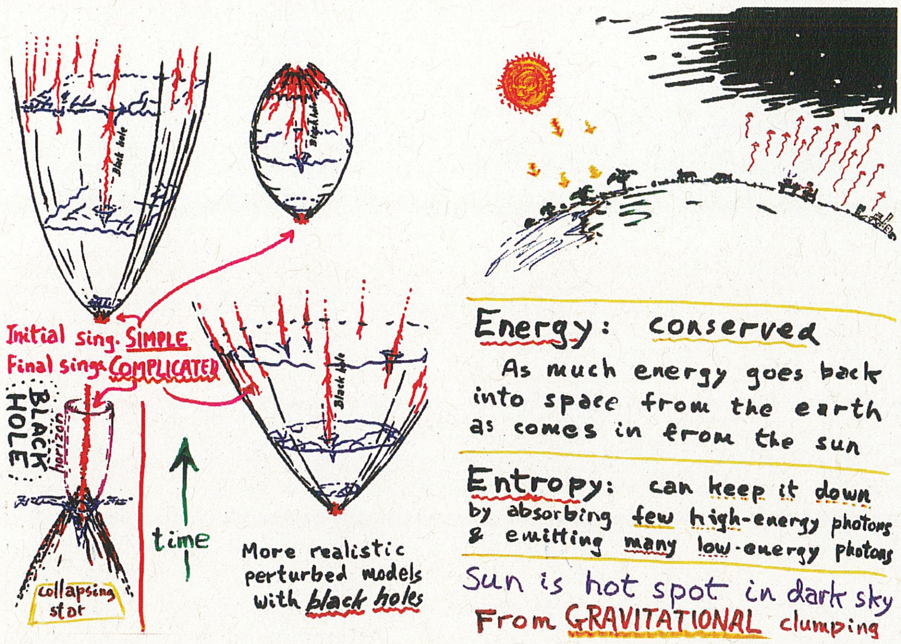

Figure 9: Roger Penrose's slides on entropy. The 'big-bang' was – as we know from the microwave background – very smooth; so it was a state of nearly maximum *microphysical* μ-entropy. But *gravitationally* it was extremely *low* g-entropy. Formation of stars (and life and civilisation etc. – all of which seems highly improbable) decreased the μ-entropy. This was possible because the decrease of μ-entropy was outweighed by the increase in g-entropy. According to Bekenstein and Hawking, most of the entropy in the Universe is in supermassive black holes.

3.3 Collisionless stellar dynamics

3.3.1 Description of collisionless particles vs. fluids: the phase-space density

- A gas – e.g. the air we breath – is very homogeneous on large scales but microscopically is 'grainy' as it is composed of particles (atoms and molecules)
- The billions of stars in an elliptical galaxy are also very smoothly distributed on large scales but the mass distribution is 'grainy' on very small scales
- At a microscopic level both can be described by specifying the positions $\mathbf{r}_i$ and velocities $\mathbf{v}_i$ for the particles
  - i being an index labelling the particles
  - and the combination $\{\mathbf{r}_i, \mathbf{v}_i\}$ – a 6 dimensional vector – being the coordinates of each particles in *phase-space*
  - note that we use velocity rather than momentum $\mathbf{p}_i = m_i \mathbf{v}_i$ (for reasons that will become apparent later)
- If we have a very large number of particles, we can usefully describe the system macroscopically by the number of particles per 6-dimensional unit volume in phase-space $f(\mathbf{r}, \mathbf{v})$
  - $f(\mathbf{r}, \mathbf{v})$ is called the *phase-space density*
- There are various different ways of thinking about $f(\mathbf{r}, \mathbf{v})$
  - in terms of the individual coordinates, $f(\mathbf{r}, \mathbf{v}) = \sum_i \delta^{(6)}(\mathbf{r} - \mathbf{r}_i, \mathbf{v} - \mathbf{v}_i)$
    * here $\delta^{(6)}$ is a 6-dimensional *Dirac delta-function*
    * integrating this over any 6-volume V gives the number of particles in that volume
  - we might also think of $f(\mathbf{r}_p, \mathbf{v}_p)$ as being the counts of objects in a grid of regular cellular volumes (6D 'pixels') with centres at positions $\mathbf{r}_p, \mathbf{v}_p$ divided by the pixel volume $(\Delta r)^3 (\Delta v)^3$
    * i.e. analogous to an image of a continuous scene described by pixel values
    * and, again if we have very large numbers of particles, it may be possible to find a 'pixel size' so that the 'root-N' statistical fluctuations of the counts in a cell are small but where the pixels are still small compared to the scale – in position or velocity – over which the density varies
    * in which case we can usefully think of $f(\mathbf{r}, \mathbf{v})$ as being some smooth function of $\mathbf{r}$ and $\mathbf{v}$

- or we can also think about making a *realisation* of a set of particles that *sample* some given phase-space density field $f(\mathbf{r}, \mathbf{v})$
  * by dividing phase-space into a very fine grid of cells and then populating each cell with a particle with probability equal to $f(\mathbf{r}, \mathbf{v})d^3rd^3v$
  * and leaving it empty with probability $1 - f(\mathbf{r}, \mathbf{v})d^3rd^3v$
  * ignoring the probability that there be more than one particle in the cell as this is an infinitesimal of higher order
  * this is called a *Poisson sample* or *Poisson realization* since the counts of objects in an finite region V will then obey Poisson statistics with mean number $\mu = \int_V d^3rd^3vf(\mathbf{r}, \mathbf{v})$
  * i.e. $P(N) = \mu^N \exp(-\mu)/N!$
- Binney and Tremaine talk about $f(\mathbf{r}, \mathbf{v})$ being analogous in a way to a quantum mechanical wave function in the sense that it it something that one can compute the evolution of, or find steady state solutions for, and from this one can compute observables such as the density – by integrating over the velocity – or velocity moments (mean velocity, velocity dispersion etc.).

3.3.2 The fluid limit

- For a *fluid* there is a more economical way of describing the system. This is because, in such a material, collisions between particles are very frequent.
  - this establishes a *locally Maxwellian distribution* of velocities
    * $p(\mathbf{v})d^3v = (2\pi\sigma^2)^{-3/2}d^3v \exp\left(-|\mathbf{v} - \overline{\mathbf{v}}|^2/2\sigma^2\right)$
  - where $\overline{\mathbf{v}}$ is the mean velocity,
  - $\sigma^2 = kT/m$ is the *velocity dispersion,*
  - and T is the temperature
- The phase-space density for a gas or fluid is then $f(\mathbf{r}, \mathbf{v}) = n(\mathbf{r})p(\mathbf{v})$ which is determined by 5 functions of position (or 'fields')
  - the *space (number) density* of particles $n(\mathbf{r})$
  - the three components of the mean velocity $\overline{\mathbf{v}}(\mathbf{r})$
  - and the temperature $T(\mathbf{r})$
- if we specify the initial values of these fields, their future evolution is constrained by
  - the equation of conservation of particles: $\partial_t n + \boldsymbol{\nabla} \cdot (n\overline{\mathbf{v}}) = 0$
  - the conservation of the three components of momentum
    * embodied in Euler's equation which states that the convective derivative of $\overline{\mathbf{v}}$ is equal to the combined pressure gradient and gravity forces
    * with pressure $P = nk_\mathrm{B}T$
- this is one short of what we need: the final fifth constraint usually being provided by an *equation of state*
  - for example that the entropy of each element of the fluid is conserved
    * which would relate the temperature to the density
    * and which provides us with the equations for sound waves in an *'adiabatic gas'*
    * though this would not work if there are shocks in the fluid – as these increase the entropy
  - of that the temperature is fixed (for an *isothermal gas* held at fixed temperature by conduction perhaps)
- For stars in a galaxy, however, collisions are completely *ineffective*. The most popular candidates for the DM such as weakly interacting massive particles (WIMPS) are also effectively collisionless and so, to a good approximation, are neutrinos.
- So we must work with the phase-space density.

3.3.3 Equations of motion

- The central assumption in galactic dynamics is that stars in a galaxy move in the Newtonian potential $\phi(\mathbf{r}, t)$ generated by the large-scale (i.e. smoothed out) density field $\rho(\mathbf{r}, t)$.
  - so ϕ and ρ are related by *Poisson's equation:*
    * $\boxed{\nabla^2 \phi = 4\pi G \rho}$
  - and the stars move according to $\mathbf{F} = m\mathbf{a}$ or
    * $\boxed{\ddot{\mathbf{r}}_i = -\boldsymbol{\nabla}\phi(\mathbf{r}_i)}$
- The latter is single equation (for each of the $3N$ components of position for N particles) that is 2nd order in time
- Alternatively, we can work with *Hamilton's equations*
  - the Hamiltonian H of a system is its total energy expressed as a function of positions, $\mathbf{r}_i$, momenta $\mathbf{p}_i$ and time.
  - Hamilton's equations are $\dot{\mathbf{r}}_i = \partial H / \partial \mathbf{p}_i$ and $\dot{\mathbf{p}}_i = -\partial H / \partial \mathbf{r}_i$
  - which are $3N$ pairs of first order coupled equations for $\mathbf{r}_i$ and $\mathbf{p}_i$ (or here $\mathbf{v}_i$):
  - $\boxed{\dot{\mathbf{r}}_i = \mathbf{v}_i}$ and $\boxed{\dot{\mathbf{v}}_i = -\boldsymbol{\nabla}\phi}$
- The physics is the same, and either way, we need to specify the same kind of initial conditions.
  - e.g. the positions and velocities $\mathbf{r}_i$ and $\mathbf{v}_i$ at some initial time
- and one can then evolve the trajectories of the particles into the future
  - this might be done in some *given* potential
  - or that generated by some given density
  - or the potential may be that which is generated by the evolving distribution of stars
    * in which case the density is given by $\rho(\mathbf{r}) = 4\pi G m n(\mathbf{r})$
    * which is what done in numerical *N-body simulations*

3.3.4 Properties of the phase-space density

There is an important distinction between the behaviour of a collisionless 'gas' of particles – such as stars in a galaxy – as pictured in 3-D space and in 6-D phase-space

- at any point (or within a small region) in *space*
  - there are generally particles moving in different directions
  - particles which are initially close together will become widely separated
  - different particles can have orbits which cross
- at any point (or within a small region) in *phase-space*
  - all the particles have the *same* (6-dimensional) velocity $\{\dot{\mathbf{r}}, \dot{\mathbf{v}}\}$
    * because $\dot{\mathbf{r}}$ and $\dot{\mathbf{v}}$ are only functions of $\mathbf{r}$ and $\mathbf{v}$
  - particles which are initially close together will stay together
  - 'orbits' of particles – i.e. the trajectories $\{\mathbf{r}(t), \mathbf{v}(t)\}$ – can never cross

This means that collisionless particles, while being quite unlike a fluid in ordinary space, actually behave as a *fluid* in phase-space. Much as a collisional gas behaves macroscopically as a fluid in ordinary 3-dimensional space.

The foregoing applies also to a collisionless *plasma*, but with the slight modification that $\dot{\mathbf{v}}$ depends on the charge-to-mass ratio of the particles. So the different components – the ions vs. the electrons say –

behave like different fluids. For gravity, as Galileo discovered, all particles have the same charge-to-mass ratio, so we have a single fluid.

It is somewhat ironic that in GR texts the matter is often modelled as an *ideal fluid*. This is probably a hold-over from the early history where the main application of GR was to relativistic stars, assumed to be composed of highly ionized plasma where collisions between particles and photons efficiently established local thermodynamic equilibrium. Outside of stellar interiors, it is essentially only for the plasma in the early universe or for very diffuse hot gas in galaxy and galaxy cluster halos where matter actually behaves like an ideal fluid.

3.3.5 The collisionless Boltzmann (or Vlasov) equation and Liouville's theorem

The phase-space density of collisionless particles actually behaves as an *incompressible* fluid. There are various ways to show this. We'll provide an illustration in 1-dimension and then give the proof for 3-dimensions.

– Liouville's theorem in 1-dimension

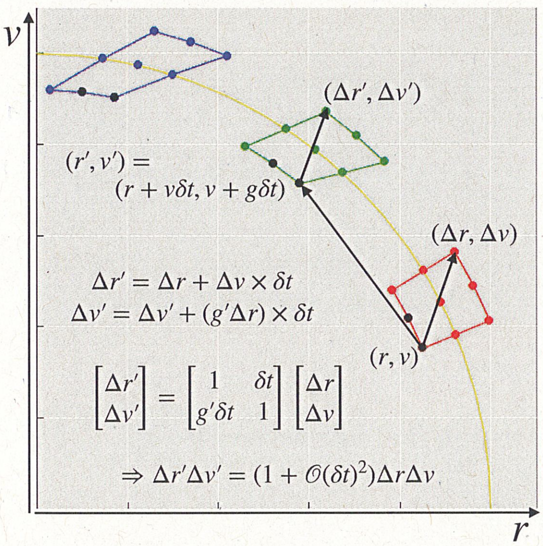

Figure 10: Liouville's theorem in 1D. A particle moves from (r, v) to (r', v') in a time interval δt under the influence of a gravitational acceleration $g(r)$ as indicated by the long arrow. Its neighbour, initially at $(r+\Delta r, v+\Delta v)$ moves to $(r'+\Delta r', v'+\Delta v')$. The new separation $(\Delta r', \Delta v')$ is related to the initial separation by a matrix multiplication. The final phase-space 'volume' – an area as we are working in 1 spatial dimension – is given by the initial volume times the determinant of the matrix. However this is equal to unity plus terms of order $(\delta t)^2$, so there is no 1st order change and hence the rate of change of the volume is constant. If we imagine a large number of particles filling this volume – whose boundaries are the red, green and blue boxes – it follows that the phase-space density of these particles must be constant in time.

- One simple way to make this plausible is to consider a 1-dimensional model, so $\{\mathbf{r}, \mathbf{v}\} \rightarrow \{r, v\}$
  - consider one particle – the 'centre particle' – with phase-space coordinates $\{r, v\}$
  - and a neighbour particles with coordinates $\{r + \Delta r, v + \Delta v\}$
  - this is at some time, which we can take to be $t = 0$
  - in short interval of time δt the centre particle moves to
    - $\{r', v'\} = \{r + v\delta t, v + g(r)\delta t\}$
  - while a neighbour particles will move to
    - $\{r' + \Delta r', v' + \Delta v'\} = \{(r + \Delta r) + (v + \Delta v)\delta t, (v + \Delta v) + g(r + \Delta r)\delta t\}$
  - hence the displacement $\{\Delta r, \Delta v\}$ of the neighbour particle becomes
    - $\{\Delta r', \Delta v'\} = \{\Delta r + \Delta v\delta t, \Delta v + \Delta r(dg/dr)\delta t\}$
  - where we have expanded $g(r + \Delta r) = g(r) + \Delta r dg/dr + \ldots$
  - which we can write as a matrix equation
    - $\begin{bmatrix}\Delta r' \\ \Delta v'\end{bmatrix} = \mathbf{M} \cdot \begin{bmatrix}\Delta r \\ \Delta v\end{bmatrix}$
  - with 2x2 matrix
    - $\mathbf{M} = \begin{bmatrix}1 & \delta t \\ (dg/dr)\delta t & 1\end{bmatrix}$

- this means that the volume in phase space (an area here, as we are working in 1 spatial dimension) defined by the 4 particles sitting on the corners of a rectangle with vertices (relative to $\{r, v\}$) of $\{0,0\}, \{\Delta r, 0\}, \{\Delta r, \Delta v\}, \{0, \Delta v\}$ evolves from $\Delta r \Delta v$ to become a distorted rectangle with area $|M| \Delta r \Delta v$ where $|\mathbf{M}|$ is the determinant of $\mathbf{M}$
- but $|\mathbf{M}| = 1 + \mathcal{O}(\delta t)^2$ (as the diagonal components are both unity)
  - so there is no first order (in δt) change in the volume
- this means that *the rate of change of the phase-space volume with time vanishes*
- and this implies that the phase-space density in the neighbourhood of the centre particles
  - which we can consider to be the number of neighbouring particles divided by the phase-space volume they occupy
- remains constant
  - or, put another way, we can say that the *convective derivative* df/dt giving the rate of change of f along the particle trajectory vanishes:
    * $\boxed{df/dt = 0}$

    which is known as *Liouville's theorem*
  - confusingly, there is another related theorem of the same name
    * the version we are considering refers to the density of particles for a single system – or a single universe, ours – in the 6-dimensional phase-space
    * the other refers to the density of *systems* – in an ensemble of systems – in 6N-dimensional space (where the configuration of all of the particles in a single system are a single point)

– **The Vlasov equation**

- A more elegant – and more general, as it applies in 3-dimensional space – approach is to start with the *equation of continuity* that expresses the conservation of particles
  - $\boxed{\partial_t f + \nabla^{(6)} \cdot (f \times \{\dot{\mathbf{r}}, \dot{\mathbf{v}}\}) = 0}$
- which says that the rate of change of the density with time in a small volume element of phase space is minus the 6-divergence obtained by 'dotting' the 6-dimensional differential operator $\nabla^{(6)} \equiv \{\nabla_\mathbf{r}, \nabla_\mathbf{v}\}$ with the 6-current $f \times \{\dot{\mathbf{r}}, \dot{\mathbf{v}}\}$
- this is the 6-D analogue of the law for conservation of mass for a fluid with mass density $\rho(\mathbf{r})$ and velocity field $\mathbf{v}(\mathbf{r})$ as illustrated in figure 11

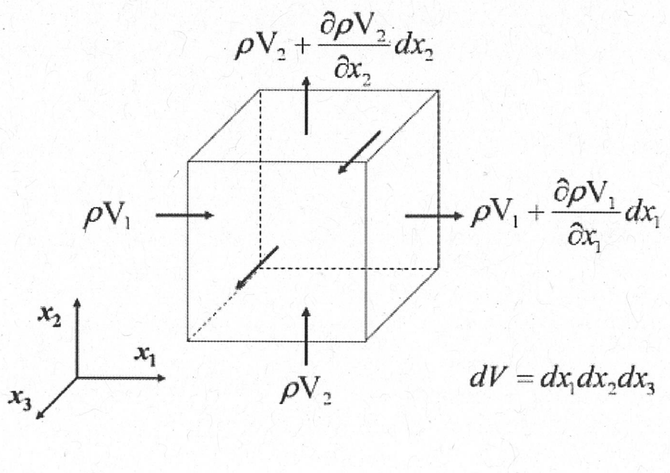

Figure 11: Conservation of mass for a fluid. The flux density is the product of the mass density $\rho(\mathbf{r})$ – though it could be the number density $n(\mathbf{r})$ if we think of the fluid as composed by a large number of identical mass particles – and the velocity flow-field $\mathbf{V}(\mathbf{r})$. The rate of change of the amount of mass in the cubical volume is obtained by differencing the flux across the 3 pairs of surfaces. Dividing by the volume gives the rate of change of ρ as minus the divergence $\nabla \cdot (\rho \mathbf{V})$ of the flux density. The continuity equation for particles in 6-D phase-space is directly analogous.

- Separating the two 3-D divergences here, this is
  - $\boxed{\partial_t f + \nabla_{\mathbf{r}} \cdot (f\dot{\mathbf{r}}) + \nabla_{\mathbf{v}} \cdot (f\dot{\mathbf{v}}) = 0}$

  and applying the chain rule, we have $\nabla_{\mathbf{r}} \cdot (f\dot{\mathbf{r}}) = \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f + f \nabla_{\mathbf{r}} \cdot \dot{\mathbf{r}} = \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f$
  - where, in the last step, we have used $\dot{\mathbf{r}} = \mathbf{v}$ and the fact that the *definition* of the partial derivative operator $\nabla_{\mathbf{r}}$ is that it gives the derivative of it's argument with respect to position holding $\mathbf{v} = \dot{\mathbf{r}}$ (and time) fixed.
- and similarly $\nabla_{\mathbf{v}} \cdot (f\dot{\mathbf{v}}) = \dot{\mathbf{v}} \cdot \nabla_{\mathbf{v}} f + f \nabla_{\mathbf{v}} \cdot \dot{\mathbf{v}} = \dot{\mathbf{v}} \cdot \nabla_{\mathbf{v}} f$
  - since, for gravity, $\dot{\mathbf{v}}$ is only a function of $\mathbf{r}$ and not $\mathbf{v}$
  - so the rate of change of $\dot{\mathbf{v}}$ with $\mathbf{v}$ at constant $\mathbf{r}$ (the definition of $\nabla_{\mathbf{v}} \dot{\mathbf{v}}$) must vanish
- Hence the continuity equation becomes
  - $\boxed{\partial_t f + \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f + \dot{\mathbf{v}} \cdot \nabla_{\mathbf{v}} f = 0}$
- which is known as the *collisionless Boltzmann equation* or as the *Vlasov equation.*
- an alternative – and more commonly travelled – route (from continuity to Vlasov) is:
  - to note that
  - $\nabla_{\mathbf{r}} \cdot (f\dot{\mathbf{r}}) + \nabla_{\mathbf{v}} \cdot (f\dot{\mathbf{v}}) = \dot{\mathbf{r}} \cdot \nabla_{\mathbf{r}} f + \dot{\mathbf{v}} \cdot \nabla_{\mathbf{v}} f + f \times [\nabla_{\mathbf{r}} \cdot \dot{\mathbf{r}} + \nabla_{\mathbf{v}} \cdot \dot{\mathbf{v}}]$
  - and to invoke Hamilton's equations $\dot{\mathbf{p}} = -\nabla_{\mathbf{r}} H(\mathbf{r}, \mathbf{p})$ and $\dot{\mathbf{r}} = \nabla_{\mathbf{p}} H(\mathbf{r}, \mathbf{p})$
  - and, since $\mathbf{p} = m\mathbf{v}$, we have $\dot{\mathbf{p}} = m\dot{\mathbf{v}}$ and $\nabla_{\mathbf{v}} \dot{\mathbf{v}} = \nabla_{\mathbf{p}} \dot{\mathbf{p}}$, so
  - $[\ldots] = \nabla_{\mathbf{r}} \cdot \dot{\mathbf{r}} + \nabla_{\mathbf{p}} \cdot \dot{\mathbf{p}} = (\nabla_{\mathbf{r}} \cdot \nabla_{\mathbf{p}} - \nabla_{\mathbf{p}} \cdot \nabla_{\mathbf{r}}) H = 0$
  - since the partial derivatives on the right hand side commute
- Either way, the LHS of the Vlasov equation is just the *convective derivative df/dt*, this being the rate of change of density f as measured by the particle moving along the trajectory $\{\mathbf{r}(t), \mathbf{v}(t)\}$ so the Vlasov equation states that
  - $\boxed{df/dt = 0}$
- *the phase-space density along any particle trajectory is constant.*

This is a powerful and general result, with vast implications.

As an aside, note that, in obtaining this, we assumed that the force on the particle was independent of velocity, as is the case for gravitational forces. But this is not the case for electromagnetic forces owing to the $q\mathbf{v} \times \mathbf{B}$ force. It turns out, however, that Liouville's theorem still applies but one cannot simply appeal to Hamilton's equations as we have done here to establish that. So in plasma physics, for example, you will see exactly the same equations used to describe the phase-space density of ions and electrons.

The unifying feature of gravitation and plasma physics here is that in both subjects it is a good approximation to ignore the effects of inter-particle interactions and consider them to be moving under the influence of some smooth force.

3.3.6 Some aspects of Liouville's theorem

– Reciprocity of volumes in real- and velocity-space

- The phase-space volume containing a collection of particles neighbouring some fiducial (or 'centre') particle does not change with time
  - so if the spatial volume decreases, the 'volume' in velocity space must increase by the same factor
  - which implies that if gravity acts so as to 'focus' particles so the density goes up, for example, the velocity dispersion must go up also (in order that the density in velocity space go down)

- this is like the adiabatic compression of a monatomic gas ($P \propto \rho^\gamma$ with adiabatic index $\gamma = 5/3$)
  - since $P = nk_BT$ and $n \propto \rho$ the above relation says that $T = P/nk_B \propto n^{2/3}$ so the typical thermal velocity v (which is $\propto \sqrt{T}$) scales as $v \propto n^{1/3}$
  - so if, for example, the size of the region in real space is halved (density going up by a factor 8) then the volume in velocity space goes up by a factor 8
- this also meshes nicely with the idea that, by virtue of the *correspondence principle* we can think of particles as being actually de Broglie waves obeying Schroedinger's equation
  - if the particles are confined in a box of changing size
  - but the wavelength of the waves scales with the box size
  - then shrinking of the spatial volume implies a decrease in wavelength and hence an increase in momemtum $p = h/\lambda$

– Particle trajectories as a mapping in phase-space

- The evolution of the phase-space density is a *mapping:* As illustrated in figure 12, if a particle with initial coordinates $\{\mathbf{r}_0, \mathbf{v}_0\}$ at $t = t_0$ moves to $\{\mathbf{r}(\{\mathbf{r}_0, \mathbf{v}_0\}), \mathbf{v}(\{\mathbf{r}_0, \mathbf{v}_0\})\}$ at time t then $f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}_0, \mathbf{v}_0, t_0)$

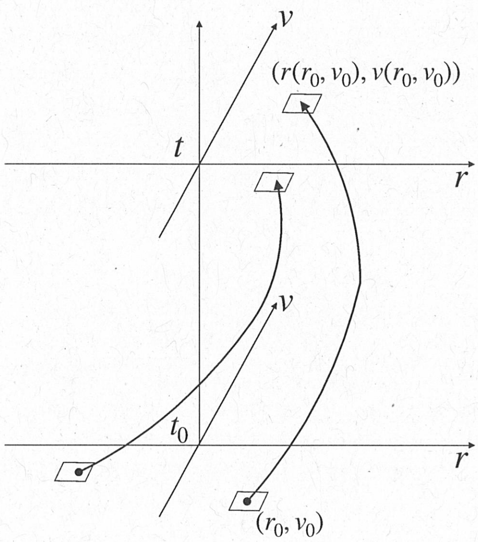

Figure 12: Illustration (in 1-D for simplicity) of the trajectories of particles moving in a gravitational potential well $\phi(\mathbf{r})$ (which may be static or evolving in time). The potential establishes a *mapping* between coordinates (r_0, v_0) in the phase-space at some time t_0 and the coordinates (r, v) at some later time. Phase-space trajectories of two particles are shown. One can imagine a 2-dimensional family of trajectories that cover the initial phase-space. The fact that $\dot{r}$ and $\dot{v}$ are functions of r and v means that these trajectories can never cross. That means that the particles behave like a fluid would (but in phase-space rather than real space) in that particles that are initially close together remain close together. The mapping under gravity has the property that it is *volume preserving*: The little parallelepiped surrounding (r_0, v_0) containing some collection of particles has the same volume as that surrounding (r, v) (and which contains the same particles). In general, these volumes will undergo shearing and rotation, but the volume containing any fixed set of particles does not change.

– Relation between Liouville's theorem and adiabatic invariance

- Phase-space density conservation is closely related to the concept of *adiabatic invariance.*
  - consider a 1-dimensional system with particles oscillating in a potential
  - let there be constant phase-space density for all orbits up to some maximum energy
  - a classic problem is "what happens to the energy of the orbits if the confining potential varies with time"
    - ∗ e.g. the motion of a pendulum bob if we slowly change the length of the pendulum
    - ∗ or, for a 2-dimensional example, what happens to the orbit of the Earth because the sun is slowly losing mass because of the solar wind
    - ∗ note that we would *not* expect the energy to be constant; that is true only in a non-varying potential

– conservation of the phase-space volume occupied by the particles implies that, for the particles on the boundary, the loop integral $\oint pdq = \oint qdp$, and which is equal to the area inside the loop, must be constant (here q is position and p is momentum) – this assumes that holes do not develop; this seems reasonable if the potential is changing slowly
– this means that, for the pendulum example, the range of positions and momenta must evolve so that $\Delta q \Delta p =$ constant
– but the energy is $E \sim \Delta p^2/m$ while the frequency is $\omega \sim v/\Delta q \propto \Delta p/\Delta q \propto \Delta p^2$ also
– hence the energy must increase in proportion to the frequency
  * i.e. consistent with $E = N\hbar\omega$ with some 'adiabatically constant' occupation number N
  * and we say that E/ω is an *adiabatic invariant*
– for the Earth's orbit the *adiabatic invariant* is, perhaps unsurprisingly, just the angular momentun

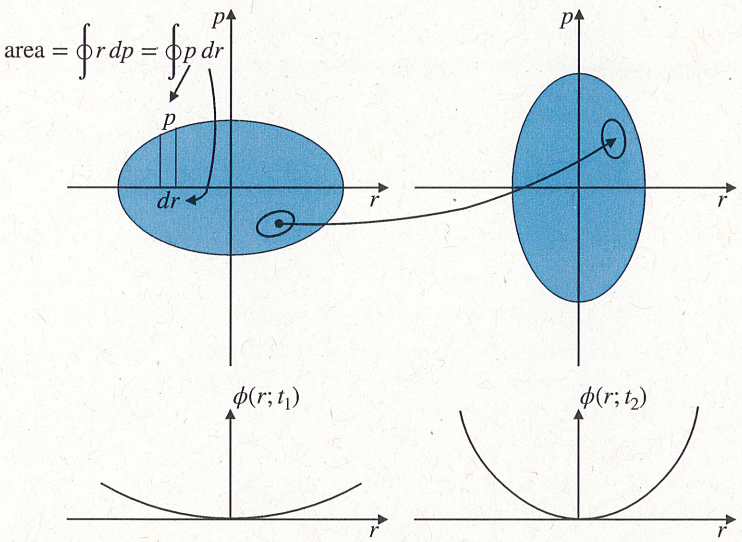

Figure 13: Adiabatic invariants are useful for considering particles that are moving in time varying potentials. The classic example is a pendulum where we slowly vary the length of the string. We do not expect energy to be conserved since the agent changing the length of the string must do or receive work. Provided the potential changes slowly, the area of phase-space within the orbit is the adiabatic invariant. This result seems inescapable if we think about an ensemble of systems that cover the region within the initial orbit and invoke Liouville's theorem. The key – an extremely useful – result is that the energy varies in proportion to the frequency.

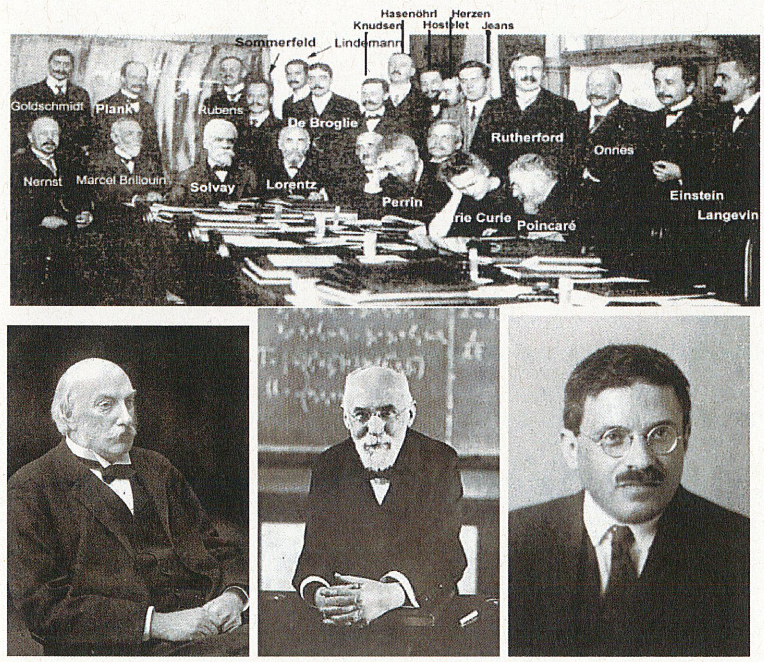

Figure 14: At the 1911 Solvay conference, Henrik Antoon Lorentz (bottom centre) posed the problem: how does the energy of a pendulum change if one slowly varies its length? The answer, unknown to him, had been given by John Strutt (bottom left), who we know as Rayleigh (he was 3rd Baron of Rayleigh), in 1902. He showed – classically – that the energy changes in proportion to the frequency. At the meeting, it was Einstein who – thinking about this quantum mechanically – gave the answer: E/ν = constant since $E = nh\nu$ and, for a continuous change the number of quanta n cannot change. It was Paul Ehrenfest (bottom right), who was not at the 1911 meeting, who developed this and emphasised the importance of adiabatic invariance in quantum mechanics.

- Liouville's theorem provides useful insight into how, for instance, particle orbits evolve in a galaxy where the potential is varying with time because matter is getting concentrated towards the centre because of cooling or decreasing because gas is getting blown out by supernovae explosions.

- Another very nice application (Dekel) is to cooling of Zel'dovich style 'pancakes' in cosmology; particles will oscillate after shell crossing. If the pancake is expanding in the two directions perpendicular to the collapse, the density of the sheet goes down and the amplitude of the oscillations will decay in a way that can be described using adiabaticity.

– **Lorentz invariance of the phase-space density**

- Another remarkable property of $f(\mathbf{r}, \mathbf{p})$ with $\mathbf{p} = \gamma m \mathbf{v}$, the relativistic momentum, is that it is *Lorentz invariant*
  - Because of Lorentz-Fitzgerald length contraction, differently moving observers will perceive different values for the spatial volume occupied d^3r occupied by a set of particles
    - if the particles are moving with respect to the observer the volume element is reduced by a factor $1/\gamma$ as compared to the volume element as measured in the rest frame
    - so the space density n is boosted by a factor γ
    - since the number of particles $n(\mathbf{r})d^3r$ is Lorentz-invariant
  - But they also perceive different volumes in momentum space d^3p
    - as shown in figure 15 d^3p is *increased* by a factor γ
  - so different observers perceive the same value for the 6-D phase space volume element d^3rd^3p
    - this combination of variables being a Lorentz invariant
  - and consequently, since $f(\mathbf{r}, \mathbf{p})d^3rd^3p$ is a number of particles, and therefore also automatically invariant, $f(\mathbf{r}, \mathbf{p})$ is a Lorentz invariant also

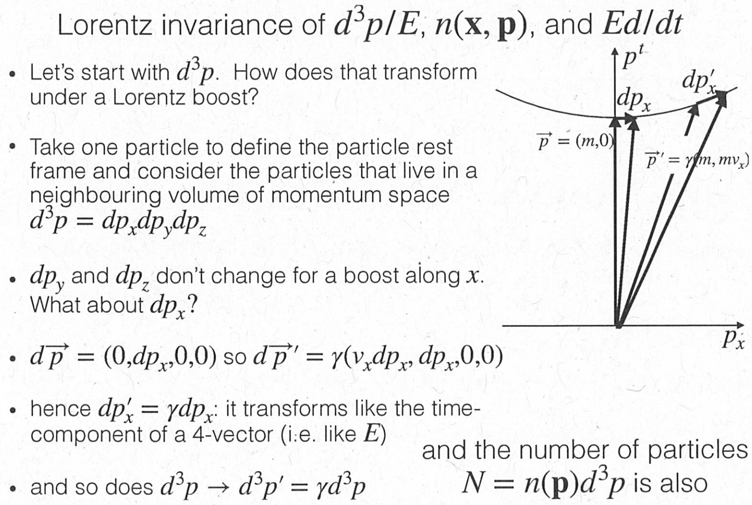

Figure 15: Lorentz transformation of real- and momentum-space densities. Consider a collection of particles at rest in some cubical volume. In the frame of a moving observer, that volume is Lorentz-Fitzgerald contracted by a factor $1/\gamma$ where, as usual, $\gamma \equiv 1/\sqrt{1 - v^2/c^2}$. That means that the space-density $n(\mathbf{r})$ is larger by a factor γ, so $n(\mathbf{r})$ transforms as the time component of a 4-vector (i.e. like the energy E of a particle). What if we have particles with a small range of momenta (in the rest-frame of one of the particles). As shown at left, in a relatively moving frame, the range of momenta is *increased*, so dp_x – assuming a boost along the x-axis – transforms like E. It follows that phase-space volume d^3pd^3x is a Lorentz invariant.

– **Planck's constant as the fundamental volume in phase-space**

- Also, again thinking of phase-space as position and momentum (rather than position and velocity), the units of phase space volume is the same as (angular momentum)^3 or, equivalently, the same as those of $\hbar^3$.
  - so constancy of phase-space density is the same as having a constant number of particles per 6-volume $\hbar^3$
  - in deriving Planck's formula for black-body radiation, we calculated the mean occupation number for the discrete allowed momentum states for waves in a box of size L. This is just the density in phase-space if we use units so $L = 1$: $\langle n(\mathbf{p}) \rangle = f(\mathbf{p})d^3p$.

– **Liouville's theorem and constancy of specific intensity**

- In optics, constancy of the specific intensity I_ν along a ray can be considered to be a manifestation of Liouville's theorem.
- but that was for static systems. More generally, if the frequency ν is varying, we have
- $\boxed{I_\nu/\nu^3 = \text{constant}}$
- to see why, recall that the definition of I_ν is that $dE = I_\nu d\nu d\Omega dA dt$ is the energy flowing in range of frequency $d\nu$ and range of direction (solid angle) $d\Omega$ through an area dA in a time dt.
- but this is also equal to
- $dE = (dA \times cdt)(fh\nu)d^3p$
- but $d^3p = p^2 dp d\Omega$ and $p = h\nu/c$ for photons so
- $I_\nu = dE/d\nu d\Omega dA dt \propto \nu^3 f$
- so if f remains constant along a ray – as Liouville tells us – then $I_\nu \propto \nu^3$
- this is very useful in cosmology where the frequency is varying
  - either because of the expansion of the universe or (in what is called the 'ISW' or 'Rees-Sciama' effect) when the photons propagate through time-varying potentials
- so we can calculate the intrinsic brightness of an object from the observed intensity if we measure the change of frequency from its redshift
- and if we have a cosmological model – with some parameters that we would like to determine – that gives us the intrinsic size from its angular size we can figure out its intrinsic luminosity
- using type 1a supernovae as 'standard-candles' this allowed astronomers in 1999 to deduce that the universe is accelerating

3.3.7 Coarse grained phase-space density

- Liouville's theorem tells us that the phase-space density does not change with time.
- In the currently popular cosmological model – the 'cold' dark matter model – the DM particles have, initially, *zero* velocity dispersion
  - i.e. their initial phase space density is *infinite*
- yet the phase-space density in the DM halo of the Milky Way is clearly finite
  - DM particles bound in the MW halo must, in order to be bound but in virial equilibrium, have velocity dispersion similar to that of stars
- The resolution of this 'paradox' is that the 'fine-grained phase-space density' is indeed effectively infinite – certainly much greater than the apparent phase-space density – but that the DM lives in an infinite density 3-dimensional 'phase-sheet' that has gotten wrapped up and folded over on itself in the process of the assembly of the galaxy
  - one might visualise the DM phase-space density as being a bit like the layers of an onion
- the apparent – or so called *coarse grained* – phase-space density obtained by averaging the density in a region of phase-space that is sliced by many sheets is much lower than the initial density
- this means that in a galaxy, or in a galaxy cluster, the apparent phase space density cannot be higher than the initial phase space density
- this was used in the classic paper of Tremaine and Gunn who showed that it was hard for massive neutrinos – a then popular candidate for DM – to constitute the DM in clusters of galaxies since the coarse-grained density was higher than that of the neutrinos in the early universe.

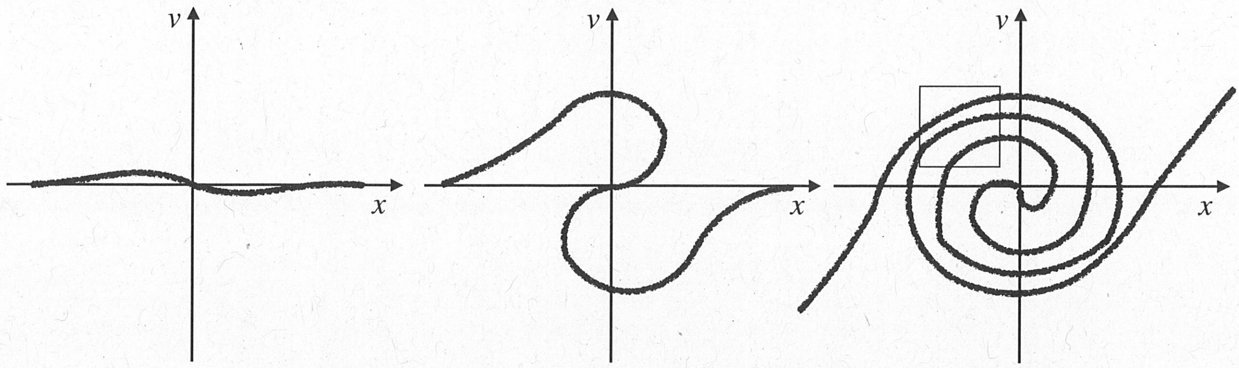

Figure 16: Coarse grained phase space density. This illustrates schematically – in 1-dimension for simplicity – how the phase-space structure of a 'dust' of initially 'cold' particles evolves as it undergoes gravitational collapse and virialisation. The left panel show the initial state where the particles at +ve x have a small negative velocity (the x-component of velocity that is). As time proceeds those particles move to smaller x and those at -ve x move in the $+x$ direction, so the density of particles increases and their velocity increases also. The effect is a 'winding up' of the 'phase-sheet'. Liouville's theorem says that the density of particles – which here have some small range of initial velocities – in the phase-sheet remains constant in time. But if we measure the average density in a box such as that shown in the right panel then what is called the *'coarse-grained phase-space density'* will be greatly decreased.

3.3.8 Violent Relaxation

- A plausible scenario for the formation of galaxies and other structures such as clusters of galaxies is that they originate as slight over-densities which eventually decouple from the overall expansion of the universe, turn around and re-collapse and *virialize*.
  - by 'virialize' we mean the process of coming to a state of *virial equilibrium* where there is neither overall contraction nor expansion of the matter and where, to a good approximation, the *virial theorem* – which states that the binding energy is minus twice the kinetic energy – holds
- in the process of collapse, bounce, and virialization the gravitational potential is changing rapidly with time
  - in fact, on a time-scale comparable to the dynamical time
- so the energies of individual particles are not constant in time (as they would be in a constant potential)
- The concept of *violent relaxation* was introduced by Donald Lynden-Bell
  - a key characteristic of this process is that it is expected to result in an object in which stars of different mass have the same distribution in *velocity* or in *energy per unit mass*
  - this is very different to the relaxation to a thermal distribution as a result of collisions in which there is equipartition of energy – so more massive particles have typical velocities that are reduced as $v \propto 1/\sqrt{m}$
- this is the reason we work with $f(\mathbf{r}, \mathbf{v})$ rather than $f(\mathbf{r}, \mathbf{p})$

3.3.9 Taking moments of the Vlasov equation

- To do stellar or galactic dynamics we couple the collisionless Boltzmann equation (or Vlasov equation) to Poisson's equation.
- This has two somewhat different types of application:

1. In the first we assume that the visible tracers make up the total mass density

6D continuity → Vlasov $\frac{df}{dt}=0$ ⇒ 3D continuity
Jeans's ⇒ Euler

- The density is then $\rho(\mathbf{r},t) = mn(\mathbf{r},t) = m\int d^3v f(\mathbf{r},\mathbf{v},t)$ – where, as mentioned m is the number weighted average mass – so we have the pair of coupled equations
  - $$\boxed{\begin{array}{c}\partial_t f + \mathbf{v}\cdot\nabla f - \nabla\phi\cdot\nabla_{\mathbf{v}} f = 0 \\ \nabla^2\phi = 4\pi G m \int d^3v f\end{array}}$$
  - where we have dropped the subscript on $\nabla_{\mathbf{r}}$
- the goal is then to find a phase space density $f(\mathbf{r},\mathbf{v},t)$ that solves these equations
  - one way to do this, in the time dependent case, is to use N-body calculations
  - in the time-independent case – as for a stable galaxy – another approach is to build a model for the potential based on the apparent distribution of stars and to try to express $f(\mathbf{r},\mathbf{v})$ as a sum over components drawn from a 'library' of orbits for that potential

2. The other approach is to drop the assumption that the visible tracers are representative of the mass density
   - this is obligatory in galaxies except in their centres
   - and use the above equations to relate observable quantities (velocities and densities) to the gravitational potential
   - we have already seen an example of this (in the 1-D case) in Oort's method to weigh the disk of the MW
     - there we simply wrote down the Jeans equation
   - next we will see how the Jeans equation in 3 spatial dimensions is obtained from continuity of phase-space density (i.e. from the Vlasov equation) by 'taking moments'
     - this means multiplying by powers of the velocity v^n and integrating $\int d^3v\ldots$
   - for $n = 0$ this gives an equation expressing conservation of particles
   - and for $n = 1$ this gives an equation expressing conservation of momentum

3.3.10 The zeroth moment: the continuity equation in 3D

- First, if we integrate the Vlasov equation over all velocities we obtain
  - $$0 = \int d^3v(\partial_t f + \mathbf{v}\cdot\nabla f - \nabla\phi\cdot\nabla_{\mathbf{v}} f)$$
    or
    $$\partial_t \int d^3v f + \nabla\cdot\int d^3v\mathbf{v}f - \nabla\phi\cdot\int d^3v\nabla_{\mathbf{v}} f = 0$$
  - where we have realised we can take ∂_t, ∇ and $\nabla\phi$ out from under the integration sign
- recognizing that the last term on the LHS integrates to give only a 'boundary term' that vanishes provided $f \to 0$ for $v \to \infty$
- and defining:
  - particle *space density*
    * $$\boxed{n(\mathbf{r},t) \equiv \int d^3v f(\mathbf{r},\mathbf{v},t)}$$
  - and the *mean velocity*
    * $$\boxed{\overline{\mathbf{v}} \equiv \int d^3v\mathbf{v}f(\mathbf{r},\mathbf{v},t) / \int d^3v f(\mathbf{r},\mathbf{v},t)}$$
- we have the (spatial) *continuity equation:*
  - $$\boxed{\partial_t n + \nabla\cdot(n\overline{\mathbf{v}}) = 0}$$
- this expresses conservation of particles in terms of number density and mean velocity (and hence *particle current* $n\overline{\mathbf{v}}$)

3.3.11 The first moment: Jeans's equation

- if we multiply the Vlasov equation by $\mathbf{v}$ and integrate over velocity we get the 3-vector equation
  - $$\boxed{\partial_t \int d^3v \mathbf{v} f + \nabla \cdot \int d^3v \mathbf{v}\mathbf{v} f - \nabla\phi \cdot \int d^3v \mathbf{v} \nabla_{\mathbf{v}} f = 0}$$
- now looking at this it is not completely obvious what, for example, the gravity $\mathbf{g} = -\nabla\phi$ is getting 'dotted' with (it is the operator $\nabla_{\mathbf{v}}$)
- things are less ambiguous if we write this using index notation
  - $$\boxed{\partial_t \int d^3v v_i f + \partial_j \int d^3v v_i v_j f - \partial_j \phi \int d^3v v_i \partial f / \partial v_j = 0}$$
  - where $\partial_i \equiv \partial / \partial_{r_i}$
  - and where we are using the 'Einstein' summation convention that any repeated index is assumed to be summed over
    - so for example we can write $\mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i$ as $a_i b_i$
- defining the velocity dispersion – or *'kinetic pressure'* – tensor
  - $$\boxed{\langle v_i v_j \rangle = \int d^3v v_i v_j f / \int d^3v f}$$

  and integrating the last term by parts gives *Jeans equation:*
  - $$\boxed{\partial_t (n\overline{v}_i) + \partial_j (n \langle v_i v_j \rangle) + n \partial_i \phi = 0}$$
- the three components of this equation express conservation (or continuity) of the 3 components of spatial momentum:
  - if we have only a single species of particle (all with the same mass) and we multiply Jeans's equation by that mass then the first term is the rate of change of i^{th} component of the momentum density $nm\overline{v}_i = \rho \overline{v}_i$
  - multiplying $\partial_t(nm\overline{v}_i)$ by a volume δV gives the rate at which the amount of momentum in that volume is changing
  - if momentum is conserved – as it is – then this must be minus the rate at which momentum is flowing out of the volume – i.e. it must be the divergence of the flux density of (the i^{th} component of the) momentum. Does this make sense?
  - the second term is the spatial gradient of $nm\langle v_i v_j \rangle = n \langle p_i v_j \rangle$ which is the *kinetic stress* or *pressure* tensor T_{ij}^{kin}, being the rate at which particles are transporting the i^{th} component of momentum in the j^{th} direction
  - so the second term is $\partial_j T_{ij}^{\text{kin}}$, which clearly *is* a divergence: it is $\nabla \cdot n \langle p_i \mathbf{v} \rangle$.
  - what about the last term $nm\partial_i \phi$? This may not look like a divergence, but it is. It is the divergence of the momentum flux density of the gravitational field:
    - $T_{ij}^{\text{grav}} = (8\pi G)^{-1}(g_i g_j - \frac{1}{2} g^2 \delta_{ij})$

    or
    - $$\boxed{\mathbf{T}_{ij}^{\text{grav}} = (8\pi G)^{-1}(\mathbf{g}\mathbf{g} - \tfrac{1}{2} g^2 \mathbf{I})}$$

    which is the gravitational analogue of the Maxwell stress tensor in electromagnetism
  - so Jeans's equation can be written as
    - $$\boxed{\partial_t(\rho \overline{\mathbf{v}}) = -\nabla \cdot \mathbf{T}}$$
    - where the left side is the rate of change of momentum density
    - and where $\mathbf{T}$ is the total stress tensor $\mathbf{T} = \mathbf{T}^{\text{kin}} + \mathbf{T}^{\text{grav}}$
      - note that since rate of change of momentum is force,

· the RHS must be a *force density*

· so this equation is the 'density' of the equation $m\mathbf{a} = \mathbf{F}$

- that was all for the special case of a single species of particle, but, as discussed previously, Jeans's equation holds for each species as long as the collisions are ineffective at transferring momentum between the different species.

- Jeans's equation is extremely powerful (as we saw in its application to the MW disk)
  - for any *equilibrium* system
    - which may be *stationary* (i.e. unmoving) so $\overline{\mathbf{v}} = 0$
    - or, more generally, may just be *static* (i.e. unchanging) so $\overline{\mathbf{v}} =$ constant
  - the first term vanishes
  - the number density of tracers n is observable, as is the velocity dispersion tensor
    - though for distant objects seen in projection we need to de-project – and that generally requires that the system be symmetric
  - so dividing by n we obtain an expression for $\nabla\phi$, and taking the divergence of this gives $4\pi G\rho$
    - and, again as already discussed, the result is independent of the choice of tracers, and, unlike the virial theorem, makes no assumption that the mass is distributed like the light
  - probably the greatest weakness of Jeans's equation is what is called the *velocity dispersion anisotropy problem:*
    - the mass density from Jeans's equation is sensitive to the *form* of the velocity dispersion tensor
    - one might be tempted to assume, for simplicity, that this tensor is isotropic
    - but if, in reality, the orbits are more radial, this will overestimate the mass

3.3.12 The Euler equation

- There is an alternative form of Jeans's equation, usually known as *Euler's equation*
  - it is useful, and rather illuminating, when applied to systems that are static (unchanging) but not stationary (unmoving)
  - an example being a *rotating* galaxy (which might be a disk or an elliptical)

! annisotropy

- to obtain this we define the dispersion of the velocity about the mean as
  - $\boxed{\sigma_{ij}^2 = \langle (v_i - \overline{v}_j)(v_j - \overline{v}_j) \rangle}$
  - so this is the dispersion of velocities relative to that of an observer 'going with the flow'
  - and, on multiplying by nm it gives the flux density of momentum in that observer's frame
- in terms of which, with a little algebra, Jeans's equation becomes
  - $(\partial_t + \overline{v}_j \partial_j)\overline{v}_i = -n^{-1}\partial_j(n\sigma_{ij}^2)) - \partial_i\phi$
- or, since the LHS is the convective derivative (the time derivative measured by our observer)
  - $\boxed{d\overline{v}_i/dt = -n^{-1}\partial_j(n\sigma_{ij}^2) - \partial_i\phi}$

$$\frac{d\underline{v}}{dt} + \frac{1}{n}\underline{\nabla}\cdot n\underline{\underline{\sigma}}^2 = 0$$

- which is usually called *Euler's equation*
- this looks very much like the equation of motion for a fluid element being subject to a pressure gradient force and a gravitational acceleration
- except that here the pressure tensor is, in general, anisotropic, and density n appearing here is not a mass density – as it would need to be in order for $\partial_j(n\sigma_{ij}^2)$ to be the real physical pressure gradient force density – but rather the space number-density of the 'tracer' particles

- and the velocity here is not, as it would be for a fluid, that of a coherent collection of particles that remain close to each other. Here the particles are passing through any volume of space with a range of directions and diverge from one another.

- One way to think of the pressure gradient term in the Euler equation in this context as giving the non-gravitational acceleration an observer would need to have to move in such a way as to be always in the local rest-frame (or 'standard of rest') of the particles that happen to be in the observer's vicinity.

- Just as we can use the other form of Jeans's equation to search for solutions that are compatible pairs of potential and phase-space distribution functions, one can use the Euler equation to find flow-fields $\mathbf{v}(\mathbf{r})$, pressure tensor fields, and potentials that are compatible. For example, one could look for stationary solutions where the flow-field might be that of a differentially rotating disk.

3.4 The collisional Boltzmann equation

- The above equations
  - one expressing conservation of particles
  - and three – for $i = 1, 2, 3$ – expressing conservation of momentum

  are valid when collisions between particles are negligible

- but in globular clusters, for instance, collisions cannot be ignored; they allow such systems to 'relax'

- more generally, if we consider two-body collisions that can scatter particles $\mathbf{v}, \mathbf{v}' \leftrightarrow \mathbf{v}'', \mathbf{v}'''$ the collisionless Boltzmann equation $df/dt = 0$ is augmented by a *collision term:* on the right hand side:
  - $$\boxed{\frac{df}{dt} = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}}$$
  - this term is (minus) the rate at which particles are being scattered out of some velocity volume element d^3v by 'reactions' $\mathbf{v}, \mathbf{v}' \rightarrow \mathbf{v}'', \mathbf{v}'''$ (see figure 17)
    * this is obtained by taking the product $f(\mathbf{v})f(\mathbf{v}')$, multiplying by a *differential cross-section area* $d\sigma(\hat{\mathbf{v}}''_c \cdot \hat{\mathbf{v}}_c)/d\Omega$ and relative velocity $|\mathbf{v}' - \mathbf{v}|$ and integrating over $\mathbf{v}'$ and $\hat{\mathbf{v}}''$
    * here the subscript c denotes a velocity in the 'centre of mass' frame, so $\mathbf{v}_c \equiv \mathbf{v} - (\mathbf{v} + \mathbf{v}')/2$ and $\mathbf{v}''_c \equiv \mathbf{v}'' - (\mathbf{v} + \mathbf{v}')/2$ etc., and the hat symbol, as usual, denotes a unit vector (i.e. the direction the particle is moving)
    * for central forces $d\sigma/d\Omega$ times $\delta\Omega$ gives the annular cross-sectional area in the 2-D impact parameter plane that, viewed in the centre of momentum frame, scatter into the directions lying in the cone of solid angle $\delta\Omega$. This is a function of the dot product between the ingoing and outgoing directions.
  - *plus* the rate at which particles are being scattered *into* d^3v by the inverse reactions $\mathbf{v}'', \mathbf{v}''' \rightarrow \mathbf{v}, \mathbf{v}'$ which involves, similarly, the product of the density of particles $f(\mathbf{v}'')f(\mathbf{v}''')$
  - so $(\partial_t f)_{\text{coll}}$ is equal to $ff' - f''f'''$ (using the notation that e.g. $f' \equiv f(\mathbf{v}')$) times factors that are independent of $f(\mathbf{v})$ (i.e. the cross-section, and relative velocity)
  - $$\boxed{df/dt = -\int d^3v' \int d^2\hat{v}''(d\sigma/d\Omega)|\mathbf{v}' - \mathbf{v}|(ff' - f''f''')}$$
  - the study of this equation and its solutions is known as *kinetic theory*

- note that we can think of this 5-dimensional integral, and calculate it if we want to, as a 9-dimensional integral over all components of $\mathbf{v}'$, $\mathbf{v}''$ and $\mathbf{v}'''$ that includes a 1-D delta-function $\delta(E + E' - E'' - E''')$ to enforce energy conservation and a 3-D delta-function $\delta^{(3)}(\mathbf{p} + \mathbf{p}' - \mathbf{p}'' - \mathbf{p}''')$ to enforce momentum conservation

- however, the fact that the rate of reactions contains the factor $ff' - f''f'''$ allows

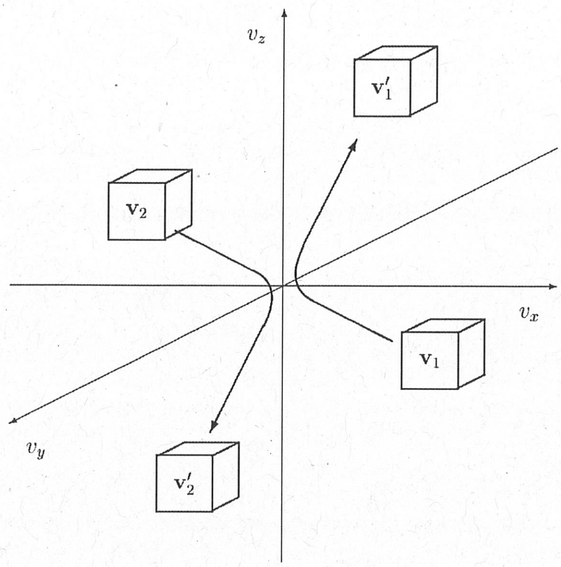

Figure 17: Illustration of a collision between two particles where initial particles are in momentum space cells labeled $\mathbf{v}_1, \mathbf{v}_2$ and end up in cells labeled $\mathbf{v}_{1'}$, $\mathbf{v}_{2'}$. In calculating the rate of change of occupation number for cell $\mathbf{v}_1$ say, we have a negative term corresponding to the 'forward' reactions as shown, but we also have a positive term arising from 'inverse' reactions, so the net rate is proportional to $-(f(\mathbf{v}_1)f(\mathbf{v}_2) - f(\mathbf{v}_{1'})f(\mathbf{v}_{2'}))$. That is for the particular combination of momenta shown. To obtain the total rate of change of f_1 we need to integrate over all possible values of $\mathbf{v}_2$ and over the direction for one of the outgoing particles (e.g. $\hat{\mathbf{v}}_{2'}$). That gives a 5-dimensional integral to perform. Several important consequences – the form for the equilibrium distribution function and Boltzmann's H-theorem – can be understood just using the fact that the net rate involves this combination of occupation numbers.

1. determine the distribution of particle velocities in a state of local equilibrium (the Maxwellian distribution)
2. and, allowing for quantum mechanical stimulation and/or blocking, obtain the thermal distribution function for bosons or fermions
3. and also understand Boltzmann's celebrated H-theorem which allows a physical explanation of entropy and the second law of thermodynamics from statistical mechanics of colliding particles

- One key property of such collisions is that *they have no effect on the continuity and Jeans (or Euler) equations.*
  - the reason is simply that collisions do not, by themselves, change the number of particles in a region of space nor the amount of 3-momentum
- One useful implication of collisions is that, provided they are efficient, they will force the phase space density to be locally Maxwellian.
  - If we consider the phase-space density $f(\mathbf{v})$ in the frame such that $\overline{v} = 0$ – the 'local standard of rest' if you like – this must be, by symmetry, spherically symmetric and therefore just some function of the energy (assuming equal mass particles)
  - it must also, in equilibrium, make $(\partial_t f)_{\rm coll} \propto f f' - f'' f'''$ vanish, or
    - $\boxed{\log f + \log f' = \log f'' + \log f'''}$
  - but, in such 2-body collisions, energy is conserved:
    - $\boxed{E + E' = E'' + E'''}$
  - and since f, in equilibrium, can only be a function of the energy, this implies
    - $\boxed{f \propto \exp(-\beta E)}$
    - with β some constant
    - which is *Boltzmann's law*
  - or, in terms of the velocity of the particles,
    - $\boxed{f \propto \exp(-\beta m v^2/2)}$
    - which is the *Maxwellian distribution*
    - β being fixed if we specify the number of particles and how much kinetic energy they have
  - it follows that the velocity dispersion tensor is *diagonal with equal diagonal elements:* $\sigma_{ij}^2 \equiv \langle (v_i - \overline{v}_i)(v_j - \overline{v}_j) \rangle = \sigma^2 \delta_{ij}$
  - where σ^2 is the velocity dispersion for each of the three components

- and which in turn is characterized by the *temperature* $T \equiv 1/k_B\beta$.
- this can only be an approximation for self-gravitating systems since particles in the tails of the distribution will escape from the system

- This is the so-called 'classical' distribution predicted for distinguishable particles
  - in reality particles such as atoms are quantum mechanical and are fundamentally indistinguishable
  - the rate of reactions, it turns out, is modified to be proportional to $ff'(1\pm f'')(1\pm f''') - f''f'''(1\pm f)(1\pm f')$ with plus sign for bosons (with the extra factors $1+f$ etc. accounting for the fact that reactions leading to already populated states are *stimulated*) and negative sign for fermions (the latter ensuring that the exclusion principle is obeyed)
  - dividing through by $(1\pm f)\ldots(1\pm f''')$ we see that the reaction rate contains a factor $f/(1\pm f) \times f'/(1\pm f') - f''/(1\pm f'') \times f'''/(1\pm f''')$
  - from which follow the 'Bose-Einstein' and 'Fermi-Dirac' equilibrium distributions
    * $\boxed{f \propto (\exp(\mu + \beta E) \mp 1)^{-1}}$
    * where the sign is + and - respectively
  - these quantum distribution functions become equivalent to the classical result for distinguishable particles in the limit that $f \ll 1$

- This is said to *close* the system of dynamical equations for a fluid or gas:
  - if we augment the 4 equations expressing conservation of particles and momentum by a 5th equation constraining the temperature
    * this might be that the system is *isothermal*
    * or *adiabatic*, in which case the entropy of any element of the fluid is the same as its initial value
  - then we have 5 equations – satisfied at all points in space – for the 5 unknown fields $n(\mathbf{r})$, $\overline{\mathbf{v}}(\mathbf{r})$ and $T(\mathbf{r})$

- the converse – unfortunately – is that in the absence of collisions we do not have closure
  - since there are 6 *degrees of freedom* in the velocity dispersion tensor
    * it being a 3x3 symmetric tensor

Now one can take the next moment of the collisional Boltzmann equation and so on. The set of equations one obtains is known as the 'BBGKY hierarchy'. Davis and Peebles applied this to galaxies in a cosmological setting.

3.4.1 Statistical mechanical entropy - Boltzmann's H-theorem

- Another fundamental result from the Boltzmann equation (with collisions) is the statistical mechanical understanding of the 2nd law of thermodynamics enshrined in *Boltzmann's H-theorem.*

- Boltzmann defined
  - $\boxed{H = -k_B \int d^3p f \log f}$
  - where k_B is the Boltzmann constant

- He showed that if the phase-space density is 'out of equilibrium'
  - i.e. the forward and backwards reactions in the collision term are not in balance

- then H will inevitably tend to increase
  - just as found for the *entropy* in classical thermodynamics

- One can understand, in broad terms, how this comes about as follows:

- the change in H under some change $\delta f(\mathbf{p})$ in $f(\mathbf{p})$ is a *functional derivative*
  - $\delta H = -k_{\rm B} \int d^3p\, \delta f \times d(f \log f)/df$
  - implying
  - $\delta H = -k_{\rm B} \int d^3p (1 + \log f) \delta f$
  - where $\delta f = (\partial f / \partial t)_{\rm coll} dt$
  - and $(\partial f / \partial t)_{\rm coll}$ is, as described above, a 5-D integral over the space of momenta that define the collisions.
- this looks rather forbidding
  - we now have a 8-D integral to perform (since we have to integrate over d^3v as well as over d^3v' and $d^2\hat{n}''$)
  - or, if you prefer, a 12-D integral with the 1+3 energy-plus-momentum conserving delta-functions
- but the desired result can be obtained rather simply as follows:
  - consider some specific 'kinematically' allowed set of momenta for the scattering $\mathbf{p}, \mathbf{p}' \leftrightarrow \mathbf{p}'', \mathbf{p}'''$ (as shown in figure 17)
  - if $ff' \neq f''f'''$ then these collisions will, *on average*, cause a change in H
  - if $ff' > f''f'''$ then collisions will tend to deplete the phase-space density around $\mathbf{p}$ and $\mathbf{p}'$ and increase the density at $\mathbf{p}''$ and $\mathbf{p}''''$
  - the contribution to the change δH is proportional to
    - $\delta H \propto [(1 + \log f) + (1 + \log f') - (1 + \log f'') - (1 + \log f''')] \times (ff' - f''f''')$
  - the 1s cancel and since $\log f + \log f' = \log(ff')$ the change in H is
    - $\boxed{\delta H \propto (\log(ff') - \log(f''f''')) \times (ff' - f''f''')}$
  - But $\log(x)$ is a monotonic function, just like x. That means the sign of the second term is the same as the sign of the first term. So H has to increase. Unless $ff' = f''f'''$, which is the condition for equilibrium.
- this was for one single kinematically allowed combination of the four 3-momenta
- to get the total change of H we need to sum over all possibilities
- but, for each of these, the change δH will still, on average, and unless the system is in equilibrium, be positive

- Boltzmann's explanation of entropy and the 2nd law of thermodynamics in terms of kinetic theory is a pinnacle of 19th century physics. At the time it was, however, controversial and strongly resisted in some quarters (among them those resistant to the idea that matter is composed of atoms!). This may have contributed to his tragic death.

- Some comments:
  - The H-theorem says that, for a system away from equilibrium, the statistical mechanical entropy will tend to increase
    - An objection to this is that it provides an 'arrow of time', while the all of the physics of the collisions is time reversible (sometimes called *Loschmidt's paradox*)
    - So a video of such a system involving collisions between particles, and for which the entropy is increasing, would, when run backwards, appear to be losing entropy.
    - To understand what is going on here, consider two streams of atoms moving in opposite directions; these will collide with each other and the velocity distribution that was initially highly anisotropic will tend to isotropise, and, provided we have sufficiently many particles, the phase space density will evolve in the same way as described by the collisional Boltzmann equation where we assume that the distribution of directions of the outgoing momenta will be distributed according to the smooth distribution $d\sigma/d\Omega$ (this function being a function of the direction cosine of the outgoing particle relative to an incoming one in the centre of momentum frame and being determined by the details of the form of the interaction). But if we were to take the final distribution of velocities with their signs reversed – to reverse the

direction of time – and feed this to the CBE this would predict a further increase of isotropy. Whereas the true time reverse would recover the initial, highly anisotropic, distribution. The reason, of course, is that in the final state there are complex correlations between the positions and velocities of the particles – which means that their impact parameters are coordinated in such a way that all the original motions are concentrated in the two original directions – and this information is lost when we simply describe the particles by $f(\mathbf{v})$.

∗ Thus there is an implicit assumption – sometimes called *the assumption of molecular chaos* or *stosszahlansätz*, and first invoked by Maxwell – that there are no specially coordinated correlations between the velocity and position of an incoming particle *and that of a particle with which it is about to collide.*

∗ But this is a perfectly reasonable assumption! The correlations described above are between a particle and another particle with which it has recently *had* a collision.

∗ So CBE is a good tool if you want to use it to approximate the evolution of colliding particles set up in a non-contrived manner. It is not a good tool for computing the result of the reverse time evolution of such particles; unless the initial state happens to be in, or close to, equilibrium, in which case the CBE will say that nothing changes.

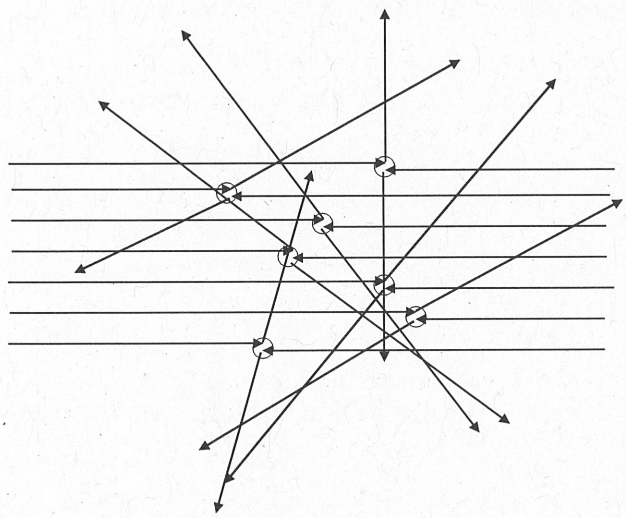

Figure 18: The assumption of 'molecular chaos' or *stosszahlansätz*. The figure illustrates the increase of Boltzmann entropy in collisions between particles in two opposing streams. In the final state, the particles are distributed over a range of momenta and the entropy – which measures, in some sense, the smoothness of the distribution function – is greatly increased. Now the time reverse of this process is also possible – as the micro-physics is time-reversible – and in the reverse process Boltzmann's theorem would be violated. But that would require carefully contrived correlations between the positions and velocities of the incoming particles. The assumption of molecular chaos asserts that there are no such correlations – though the positions and velocities of *outgoing* particles would, of course, be highly correlated.

- The equilibrium distribution – the Boltzmann distribution for classical particles – is that distribution which maximises the entropy. This is calculated by requiring stationarity of the entropy H with respect to a variation δf with constraints such as number of particles and total energy being added with Lagrange multipliers

- One can of course obtain the form of the entropy by considering (and maximising) the logarithm of W. This being the number of ways of distributing particles to cells (or 'bins') with discretized energies E_i.

- If there are n_i particles in the ith bin, one has $W = N!/\prod_i n_i!$ so $\log W = \log(N!) - \sum_i \log(n_i!)$. Assuming these occupation numbers are large one can use Stirlings formula: $\log(n!) \simeq n \log n$. So $\log W = \text{constant} - \sum_i n_i \log(n_i)$.

- This is equivalent to Boltzmann's entropy, and the content of the H-theorem is the reasonable claim that systems will evolve towards states for which $\log W$ is large.

- What this does not give us, which the CBE does, is a way to calculate, for example, the time it would take for equilibrium to be reached.

3.4.2 Application of the H-theorem to gravitating systems

- Boltzmann's $H = -\int d^3v f(\mathbf{v}) \log(f(\mathbf{v}))$ is the entropy per unit volume for a locally homogeneous collisional gas (we're dropping the constant k_B).
- multiplying by a volume element gives the total entropy S in that volume
- while the total number of particles is $N = V \int d^3v f(\mathbf{v})$
- dividing the former by the latter gives the entropy per particle (or the *specific entropy*):
  - $\boxed{s = -\int d^3v f(\mathbf{v}) \log(f(\mathbf{v})) / \int d^3v f(\mathbf{v})}$
- So the mean entropy per particle is $\langle -\log f \rangle_\mathbf{v} = \langle \log f^{-1} \rangle_\mathbf{v}$.
- The total entropy S of a self-gravitating system will depend on the details of the shape – in position and velocity – of the system.
- But if we consider a *family* of similar systems – similar in the sense that they have the same 6-dimensional shape and differ only in the overall extent in physical size r and velocity distribution width v – then these have $f^{-1} \sim r^3 v^3/N$ which is proportional to N^2/v^3 (since $v \sim \sqrt{GmN/r} \Rightarrow r \propto N/v^2 \Rightarrow f^{-1} \propto N^2/v^3$)
- the specific entropy is $s = \text{constant} + \log N^2 + \log T^{-3/2}$ where $T \equiv v^2$ is the *kinetic temperature.*
- if we keep the number of particles in the system fixed, but add or subtract energy (and thus decrease or increase T - recalling the negativity of specific heat), then $s = \text{constant} + \log T^{-3/2}$ and the total entropy is $S = \text{constant} + N \log T^{-3/2}$
  - thus the colder (or larger) the system the higher the entropy
  - and the change in entropy with temperature is $dS = -1.5 N dT/T$
- Consider two similar systems: one with N_1 particles and $T = T_1$ and the other with N_2 and T_2 that somehow exchange some energy dE that flows from system 1 to system 2 while keeping the numbers fixed.
- The change in the entropy of system 1 is $dS_1 = -1.5 N_1 dT_1/T_1 = -1.5 dE/T_1$ while $dS_2 = -1.5 N_1 dT_2/T_2 = +1.5 dE/T_2$ so the change in the total entropy is
  - $\boxed{dS = dS_1 + dS_2 = 1.5 dE(1/T_2 - 1/T_1)}$
- This will be positive (for positive dE) if $T_1 > T_2$.
- Thus for the total entropy to increase requires that energy flow from the hotter to the cooler body (as assumed without proof earlier).

3.4.3 Maximum entropy image reconstruction

- Another (completely unrelated) application of Boltzmann's statistical mechanical entropy formula in astrophysics is *maximum entropy image reconstruction*
  - astronomers are often provided with some kind of 'projection'
  - e.g. we might have a 2D image of the projected density of matter in an object from which we would like to infer the 3-dimensional structure
  - or we might have a light-curve of a tumbling asteroid illuminated by the sun whose shape we would like to infer
  - such reconstruction problems are usually somewhat 'ill-posed', with no unique solution
  - in such problems it can be very useful to determine that solution – for a 3-D 'image' $f(\mathbf{r})$ say – that maximizes the analogue of Boltzmann's entropy $-\int d^3r f(\mathbf{r}) \log f(\mathbf{r})$

– this is sometimes described as 'regularization'
– some justification for this comes from either
  * the idea that the entropy measures the probability to have a particular set of occupation numbers
  * or that the maximum entropy image is, in some sense, the 'smoothest' image that is compatible with the data

- MAXENT, as it is known, is an example of a 'Bayesian' approach to solving statistical problems

BAYESIAN MAXIMUM ENTROPY IMAGE RECONSTRUCTION

John Skilling
Dept. of Applied Mathematics and Theoretical Physics
Silver Street
Cambridge CB3 9EW, U.K.

Stephen F. Gull
Cavendish Laboratory
Madingley Road
Cambridge CB3 0HE, U.K.

ABSTRACT

This paper presents a Bayesian interpretation of maximum entropy image reconstruction and shows that $\exp(\alpha S(f,m))$, where $S(f,m)$ is the entropy of image f relative to model m, is the only consistent prior probability distribution for positive, additive images. It also leads to a natural choice for the regularizing parameter α, that supersedes the traditional practice of setting $\chi^2 = N$. The new condition is that the dimensionless measure of structure $-2\alpha S$ should be equal to the number of good singular values contained in the data. The performance of this new condition is discussed with reference to image deconvolution, but leads to a reconstruction that is visually disappointing. A deeper hypothesis space is proposed that overcomes these difficulties, by allowing for spatial correlations across the image.

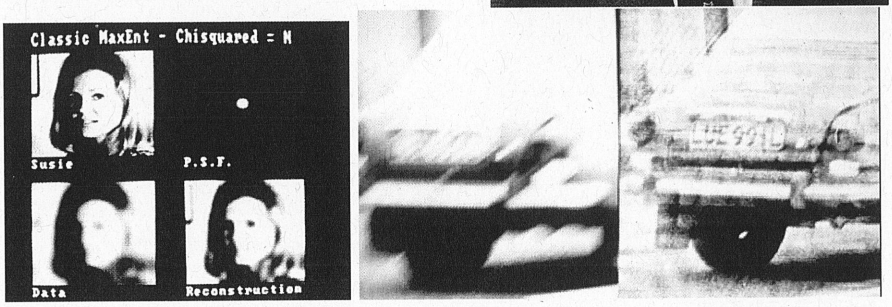

Figure 19: MAXENT image reconstruction. Upper left is the abstract from the paper by Skilling and Gull proposing the method with example shown below. Upper right is Ed Jaynes; who is one of the founding fathers of MAXENT in particular and Bayesian techniques in general. Lower right is an example of image reconstruction applied to 'de-blur' an image. A simple model for MAXENT is the 'roomful of monkeys', who we image to generate images at random by tossing marbles onto a floor covered in egg crates. The MAXENT image is, of all the images consistent with the data, the one that occurs most frequently. Techniques of this kind are extremely powerful in e.g. reconstructing images from 'visibilities' in radio interferometry.

4 Galaxy evolution

Galaxy evolution is a complex and 'fact-rich' subject that comprises

- evolution of the 'demographics' revealed from the luminosity function
- 'archeological' measurements of the history of star formation from population synthesis modelling
- observations of merging of galaxies and untangling of the complex interplay between galaxies, merging and 'active galactic nuclei' (AGN)

In the popular model for structure formation from 'cold-dark-matter' halos of galaxies assemble hierarchically with the first haloes being of low mass forming at $z \sim 30$ and with larger halos being built up later by merging.

In this picture, it was realised early on that cooling of gas would not be effective in the most massive halos forming recently (galaxy clusters) and that gas physics explained very crudely why there was an upper limit to the luminosity of galaxies (and why, for instance, the $10^{15} M_\odot$ coma cluster is a cluster of a few hundreds of bright galaxies rather than one single ultra-luminous galaxy).

But observations have shown that the situation is considerably more complex than this.

For one thing, measurements of the luminosity function show that the bright – predominantly elliptical – galaxies were already in place by $z \simeq 1$ and it is more modest – sub-L_* galaxies – that have been appearing more recently. This is known as 'down-sizing'.

Another interesting discovery was so-called 'ultra-luminous infra-red galaxies' (ULIRGS) that had been completely missed in optical surveys as the visible light was shrouded by dust and was being 're-processed' into the IR bands, so all of the luminosity of a bright galaxy was being emitted in the IR.

To confuse matters further, X-ray and other observations showed that these galaxies also often hosted AGN.

The picture that has emerged is that of a cycle of evolution in which a pair of galaxies merge and, in the process, gas is able to lose angular momentum and become dense and form copious amounts of stars and dust (which absorbs and re-processes the optical starlight). This is known as the 'star-burst' phase. At the same time, gas is also funnelled onto the supermassive black hole at the centre of the newly formed galaxy, adding to the luminosity and eventually 'quenching' the star formation and, after the clearing out of the dust, resulting in a 'red-and-dead' elliptical galaxy, perhaps with an active nucleus at the centre. This meshes well with many observational facts, including the observations of radio emission from jets emerging from AGN in elliptical galaxies. It can also be reproduced plausibly in numerical hydro-dynamical simulations. But there are a lot of details to be worked out such as whether the rate of mergers actually matches what is needed.

Central to this paradigm is the 'unified model' for AGN where the BH and accretion disk are surrounded by a dense and optically thick torus and that the different classes of active galaxies we see is fundamentally the result of looking at a single population but from different angles. In this model, in a face-on galaxy we see the BH and high-velocity gas in the accretion disk directly – the so called broad-line region – while in edge-on systems we only see the nuclear light scattered by lower-velocity clouds above and below the plane of the torus – the 'narrow-line regions'.