400A - Virial theorem

Mathieu Renzo

September 12, 2025

Materials: Chapter 2 of Onno Pols' lecture notes, Chapter 3 of Kippenhahn's book, Sec. 1.5 of "Tapestry of Modern Astrophysics" by S. N. Shore, pages 134-139 Clayton's book.

Virial theorem

In the previous lecture, we have discussed that it is appropriate to use thermodynamics to describe the *local* properties of the stellar gas. We had already established that (self-)gravity determines the pressure gradient in the star through the hydrostatic equilibrium equation (which is just the particular form that momentum conservation takes in a stellar context). Therefore, in this lecture we will explore the relation there may be between gravitational physics and thermodynamics which will allow us to get some crucial facts about stars.

The virial theorem relates the kinetic and potential energy of system of point masses interacting through long-range forces, and thus allows one to **connect thermodynamics and gravitational physics**. It is a statistical statement on mutually interacting particles (e.g., because of their mutual gravity!), and thus it is very general. The etymology of the name comes from "vir" which is latin for "forces".

(Almost fully) General derivation

Let's start with a very general derivation (see for example Goldstein's classical mechanics book or Shore's Tapesty of Modern Astrophysics Sec. 1.5). For a collection of particles with (constant) mass m_i, Netwon's second law states that the change in their momentum is equal to the total force they are subject to:

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i \ . \tag{1}$$

Note that here $\mathbf{r}_i$ and $\mathbf{F}_i$ are vector quantities. We take the *first-order spatial moment of this distribution*, which means we take the dot product of each term by the position of the particles and sum over all the particles. Why? The reason is to make *something like the work done by forces (and thus the potential energy) appear in the second term*:

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \mathbf{r}_i = \sum_i \mathbf{F}_i \cdot \mathbf{r}_i \ . \tag{2}$$

The r.h.s. looks like a sum of work done by the total foce on each particle: $\sum_i \mathbf{F}_i \cdot \mathbf{r}_i \equiv W$. This is the so called "virial", which contains only long-range forces between the particles (collisions cancel out in the sum because of Netwon's 3^{rd} law). This by conservation of energy has to be equal to the net potential energy available E_{pot}.

The l.h.s. can be re-written using the chain rule for the second time-derivative:

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \mathbf{r}_i = \sum_i \frac{1}{2} m_i \frac{d^2}{dt^2} r_i^2 - \sum_i m_i \cdot \dot{r_i}^2 \tag{3}$$

And now we have on the right hand side (assuming the mass of the particles to be constant) twice the second time derivative of the moment of inertia $I = \sum_i m_i r_i^2$ and twice the kinetic energy $E_{\mathrm{kin}} = \sum_i m_i v_i^2/2$ (because of course $v_i = \dot{r_i}$). So putting the r.h.s. of this last equation in the first spatial moment of Newton's second law (Eq. 2) we obtain:

$$2\mathrm{E}_{\mathrm{kin}} + \mathrm{E}_{\mathrm{pot}} = \frac{1}{2}\ddot{I} \tag{4}$$

This is the general form of the Virial theorem derived by Clausius. In general, in the virial theorem there is a second derivative of the moment of inertia on the r.h.s., which is often forgotten. We did not specify almost anything about the system (see Goldstein for a momentum-based derivation where the assumption of constant masses $\mathrm{m_i}$ of the particles implicit in the way we wrote Newton's second law is not even needed).

The virial theorem effectively says that *for a bound system of self-interacting particles*, *there is a limited amount of potential energy that can be extracted*, because part of the potential will need to go into the kinetic terms and the acceleration of the moment of inertia.

So far, this general form doesn't give us much intuition on its meaning. We have a relation between the kinetic energy of a collection of particles, its potential energy, and the second time-derivative of its moment of inertia. But in the frame of the center of mass of such collection, the kinetic energy of the particles is a measure of their thermal energy, and this is non-trivial: why would the thermal energy know about the potential energy! This statement is however very general and its implication for stellar structure and evolution are profound.

Derivation for binary orbits

As a demonstration for the generality of the virial theorem, we can apply it to a collection of particle with only 2 particles subject only to the gravitational force, that is a binary system of two point-mass stars on a Keplerian orbit!

One can write the total energy of the binary as:

$$E_{\mathrm{tot}} = \mathrm{E}_{\mathrm{kin1}} + \mathrm{E}_{\mathrm{kin2}} + \mathrm{E}_{\mathrm{pot}} \equiv \mathrm{E}_{\mathrm{kin}} + \mathrm{E}_{\mathrm{pot}} \equiv -\frac{GM_1M_2}{2a} \quad , \tag{5}$$

where in the last term we are using Kepler's laws. The gravitational potential energy is $E_{\mathrm{pot}} = -GM_1M_2/a$, thus making it explicit we obtain: $E_{kin} = |E_{pot}|/2$. Since Keplerian orbits in Newtonian gravity are stable (energy losses through gravitational waves make this statement not physically true, but this is only relevant for close binaries of compact objects), this implies that the second (and first) time-derivative of the moment of inertia of the orbit is zero, and thus this *is* the virial theorem! *Keplerian binary orbits are "virialized".*

N.B.: Here we considered the case of just 2 particles, but nothing stops us to consider $i = 2...N$ with N that can reach $\sim 10^6$ and apply the virial theorem to *clusters* of stars. Indeed, the virial theorem is useful to initialize the positions and velocities of stars in a simulated self-gravitating cluster such as globular clusters (the question usually is whether the second-time derivative of the cluster's moment of inertia can or cannot be approximated with zero).

Derivation for stars

Let's now re-derive the Virial theorem for a star. Let's assume a continuous description of the stellar gas, hydrostatic equilibrium and spherical symmetry, and write explicitly the forces acting on each shell dr of gas (cf. hydrostatic momentum balance seen before):

$$\frac{dP}{dm} = -\frac{Gm}{4\pi r^4} \tag{6}$$

By analogy with what is done in the general derivation above, we want to make the gravitational potential appear on the r.h.s., which can be done by multiplying by $4\pi r^3$ and integrating in dm over the entire mass of the star (from 0 to M).

$$\int_0^M -\frac{Gm}{4\pi r^4} \times 4\pi r^3 dm = \int_0^M -\frac{Gm}{r} dm \equiv -\mathrm{E}_{\mathrm{pot}} \quad . \tag{7}$$

(so that $E_{\text{pot}} = |\int_0^M -Gm/rdm| > 0$). The l.h.s. becomes integrating per parts

$$\int_0^M \frac{dP}{dm} 4\pi r^3 dm = \left[4\pi r^3 P\right]_0^M - 3\int_0^M 4\pi r^2 \frac{\partial r}{\partial m} P dm \quad , \tag{8}$$

where the first term is zero: $P(M) = 0$ at the outer surface of the star, and in the center $r = 0$ by definition. For the second term, we can use the mass continuity equation and obtain $-3\int_0^M P/\rho dm$, thus, putting back the pieces together:

$$\int_0^M \frac{Gmdm}{r} = 3\int_0^M \frac{P}{\rho} dm \quad . \tag{9}$$

Note that if we had kept a non-zero $\rho\ddot{\mathbf{r}}$ term in Eq. 6, we would again obtain a term depending on the moment of inertia and the bulk kinetic energy of the stellar gas (see for example sec. 4.2.1 in Shore's "Tapestry of modern astrophysics"). From Eq. 7, we have already interpreted the l.h.s. of Eq. 9 as the gravitational potential energy normalized to zero at infinity.

Let's interpret the r.h.s., which in the general derivation would be part of the $\sum_i \mathbf{F}\cdot\mathbf{r}_i$ term. By dimensional analysis we know it has to have the dimension of an energy. In fact, still by dimensional analysis we can infer that [P/ρ] = [E]/[L^3] / [M]/[L^3] = [E]/[M] has the units of a specific energy (i.e., energy per unit mass), and thus $P/\rho \propto u$ with $u \equiv u_{\mathrm{int}}$ specific internal energy. Let's rewrite this as $u = \Phi P/\rho$ with Φ unknown dimensionless constant.

To specify Φ, let's consider the second law of thermodynamics per unit mass: *the heat exchanged by a gas* (dq) *is equal to the change in internal energy* (du) *plus the work done* (Pdv, with v the specific volume, i.e. the volume per unit mass). We can use this to *define* the specific entropy s, and also re-write this in terms of density instead of specific volume \$v\$=(1 mass unit)$/\rho \Rightarrow dv = -d\rho/\rho^2$:

$$dq = Tds = du + Pdv = du - \frac{P}{\rho^2} d\rho \quad , \tag{10}$$

We want to derive a relation between u, P, and ρ. We are considering the momentum conservation (Eq. 6), so something related to dynamics (which acts fast compared to thermal processes – you can verify this comparing timescales after the end of this lecture!), therefore let's consider an adiabatic process where by definition there is no heat exchange. Thus, $dq = 0$ and $du = P/\rho^2 d\rho$.

Now by differentiating $u = \Phi P/\rho$ we get $du = \Phi(dP/\rho - P/\rho^2 d\rho) \equiv P/\rho^2 d\rho$ (where we use Eq. 10 with $dq = 0$ in the last step), that can be re-arranged into $\Phi/(1+\Phi)dP/P = d\rho/\rho$, or in other words $(1+\Phi)/\Phi = d\log(P)/d\log(\rho) \equiv \Gamma_1$ because the derivative has to be taken assuming no heat exchange, that is at constant entropy, so that is by definition the first adiabatic index Γ_1. Ultimately, we can put all this together and write

$$\frac{P}{\rho} = (\Gamma - 1)u \quad , \tag{11}$$

where $\Gamma \equiv \Gamma_1$ is the first adiabatic index.

Finally, substituting in the first-order moment of the hydrostatic equilibrium Eq. 9 equation we have:

$$3\int_0^M \frac{P}{\rho} dm = 3\int_0^M (\Gamma - 1)udm \quad . \tag{12}$$

Assuming that Γ is constant throughout the star, we can take the parenthesis out of the integral. We can define $E_{\mathrm{int}} = \int_0^M udm$ and rewrite the above as $3(\Gamma - 1)E - \mathrm{int}$, and thus

$$-E_{\mathrm{pot}} = 3(\Gamma - 1)E_{\mathrm{int}} \tag{13}$$

For a monoatomic gas where each particle has 3 degrees of freedom (motion in 3 direction, no internal degrees of freedom for point-particles), $\Gamma = 5/3$, and we obtain $|E_{\mathrm{pot}}| = 2E_{\mathrm{int}}$ where the internal energy is due to the thermal motion of gas, in agreement with the general form assuming hydro-*static* equilibrium ($\Rightarrow d^2I/dt^2 = 0$).

Now lets define the total energy of the gas: $E_{\mathrm{tot}} = E_{\mathrm{pot}} + E_{\mathrm{int}}$. For a star to be bound, $E_{\mathrm{tot}} < 0$. Because of the virial theorem in Eq. 13, we can rewrite this as

$$\mathrm{E_{tot}} = \mathrm{E_{pot}}\frac{3\Gamma - 4}{3(\Gamma - 1)} \equiv -(3\Gamma - 4)\mathrm{E_{int}} \quad . \tag{14}$$

Again, for a monoatomic gas with $\Gamma = 5/3$ we recover $E_{\mathrm{tot}} = E_{\mathrm{pot}}/2$, the total energy is half of the gravitational binding energy!

Stability criterion

From the form of $E_{\mathrm{tot}} \equiv E_{\mathrm{tot}}(E_{\mathrm{int}})$, since E_{int} is a quantity that is always non-negative by definition (think thermal energy!), we see that for the star to be bound, that is $E_{\mathrm{tot}} < 0$, then a necessary condition is $\Gamma > 4/3$. One can immediately see that if $\Gamma < 4/3$, then E_{tot} changes sign. If $E_{\mathrm{tot}} > 0$ the stellar gas is unbound: the kinetic energy due to thermal motion is sufficient to overcome the gravitational potential and the gas will fly out.

One can encounter cases in stellar evolution where $\Gamma \leq 4/3$ (e.g., because of recombination that changes the mean molecular weight or pair-production). In general though Γ is *not* constant throughout the star as we assumed to pull the $\Gamma - 1$ factor out of the integral. In physical situations where $\Gamma \leq 4/3$ this typically occurs first *locally* somewhere in the star, and there can be a *local* rearrangement of the stellar gas that prevents catastrophic consequences. However, this local rearrangements are not always sufficient or even possible, and sometimes *global* consequences of the impossibility of stability occur: for example the theorized (pulsational) pair instability supernovae where a (very) massive star is completely obliterated and leaves behind no black hole.

- **Q**: for a star made only of photons, what is Γ? **Hint**: think of the $P \equiv P(u)$ relation!

Use of gravitational energy in a star: negative heat capacity

Let's consider a star made of a perfect mono-atomic gas with $\Gamma = 5/3$. Let's say that for some reason this star is contracting (e.g., because it is in its formation process). Let's assume this contraction is quasi-static, meaning that at any point in time the hydrostatic equilibrium assumption holds, and all the bulk motion of the gas is very slow compared to thermal velocities, that is $v \ll v_{th} = (2k_BT/m)^{1/2}$, and thus also highly sub-sonic, $v \ll c_{\mathrm{sound}}$.

The contraction increases the gravitational potential energy $E_{\mathrm{pot}} \propto GM^2/R$ since M is by assumption constant and R decreases. The total energy E_{tot} also increases, but only by half the amount that the gravitational potential increases. The other half goes into internal thermal energy of the gas because of the Virial theorem! This is the limit in the amount of work that can be extracted set by the virial theorem mentioned above.

For an ideal gas, the internal energy is related to the mean temperature by

$$E_{\mathrm{int}} = \frac{\#\text{degrees of freedom}}{2} N k_B \langle T \rangle \tag{15}$$

where $N = \int \rho/(\mu m_u) dV = M/(\mu m_u)$ is the number of particles, with μ mean molecular weight assumed to be constant, $\langle T \rangle$ is the average temperature in the star, and the number of degrees of freedom is 3 for a monoatomic gas. From the Virial theorem it follows that:

$$E_{\mathrm{int}} = -\frac{E_{\mathrm{pot}}}{2} \Rightarrow \frac{3}{2}\frac{M}{\mu m_u} k_B \langle T \rangle = C\frac{GM^2}{R} \quad , \tag{16}$$

where $C = \{\int_0^M Gmdm/r\}/\{GM^2/R\}$ is a constant of $\mathcal{O}(1)$ that depends on the mass distribution in the star. Thus:

$$\langle T \rangle = \frac{2\mu m_u}{3k_B} C \frac{GM}{R} \propto \frac{\mu M}{R} \quad . \tag{17}$$

From this equation, several important facts follow for any self-gravitating star in hydrostatic equilibrium:

- the mean temperature of a star depends only on its mass M and radius R (and chemical composition through μ);
- *as a star contracts* (R decreases at constant mass M), *the temperature must rise*! This, as we will see, governs the evolution of stars.
- A self-gravitating collection of particle with finite temperature must radiate away energy, thus it will lose energy. This energy loss, since $E_{\mathrm{tot}} = |E_{\mathrm{pot}}|/2 = -CGM^2/2R < 0$ implies that R must decrease. But then, $\langle T \rangle$ must increase! This is the "gravothermal" collapse of a cloud/star. As thermal energy is lost to radiation at the surface, the (average) temperature raises! This is why stars can be thought of objects with *negative* heat capacity: it heats up as it loses energy! This is a property typical of self-gravitating systems only (stars and gravitationally bound stellar clusters) and because of the virial theorem it does not violate energy conservation.
- The gravothermal collapse must go on until either:
  1. an internal energy source, compensating for the surface energy loss kicks in, or
  2. the ideal gas approximation does not hold anymore.

As we will see, both 1. and 2. occur in nature: 1. is the typical option for stars using nuclear fusion as internal energy source to delay the collapse, and 2. is what occurs for white dwarfs, where quantum mechanical effects stop the collapse.

Thus, *gravity determines not only the structure (through the hydrostatic equilibrium equation) but also the evolution of a star*. It dictates that as the star loses energy, it must heat up. Because of this, at some point nuclear fusion can occur (as we will discuss later): *stars don't shine because they burn, stars burn because they shine*. Without the loss of (internal/thermal) energy by radiation they would not contract, without contracting they would not reach temperatures high enough to do nuclear fusion. **The nuclear fusion is a consequence of the fact that stars shine, and not the cause**. Stars shine as any object with finite temperature must do.

Internal energy sources such as nuclear fusion ultimately only delay the gravothermal collapse of the stars until either the ideal gas approximation does not hold (for white dwarfs) or even nuclear fusion cannot stop the collapse, and gravity wins, leading to a supernova explosion and/or the formation of a black hole.

Kelvin-Helmholtz timescale

This is by definition the timescale it takes a star to radiate away all its internal energy at a constant rate in absence of any other energy sources. Note that this is a *global* timescale for the whole star!

Let's call L the "luminosity of the star", that is the rate at which is loses energy from radiating away photons at its surface. Note that L has the units of power [L] = [E]/[t].

Then by definition:

$$\tau_{\mathrm{KH}} = \frac{\mathrm{E_{int}}}{L} \equiv \frac{\mathrm{E_{pot}}}{2L} \simeq \frac{GM^2}{2RL} \quad , \tag{18}$$

where we have used the virial theorem and set $C \simeq 1$. We can scale all the quantities to Solar values and obtain:

$$\tau_{KH} \simeq 1.57 \times 10^7 \mathrm{years} \times \left(\frac{M}{M_\odot}\right)^2 \left(\frac{R}{R_\odot}\right)^{-1} \left(\frac{L}{L_\odot}\right)^{-1} \quad . \tag{19}$$

Clearly $\tau_{KH} \gg \tau_{ff}$, and also $\gg$ {human timescales}: it is hard to get direct observational evidence that stars are in thermal equilibrium. In fact, the name of this timescale comes from proposals in the late 19$^{\mathrm{th}}$ century by Kelvin and Helmholtz that the Sun may be out of thermal equilibrium and contracting, meaning it would have a lifetime of the order of ~ 10 million years – this was in contrast with geological evidence (and with the timescale necessary for Darwinian evolution), leading to a great debate that was ultimately settled with the discovery of nuclear energy as a potential source of energy in the 1930s (by primarily Hans Bethe and collaborators), see G. Shaviv "Life of stars" for a detailed discussion.

- **Q**: A star may lose energy also by means other than photons, especially neutrinos. If such energy losses are important, how can we modify Eq. 18?

Homework

Using the virial theorem:

- Estimate the average temperature of the Sun and compare it with its surface temperature. Is the Sun in global thermal equilibrium? (you can assume $C \sim 1$ in the notation used in class/in my notes, or use the mass and radius profile from the `MESA-web` model you already computed to calculate a more precise value for C)

- Find an order of magnitude ($\sim$) relation between the average sound speed in the star and the escape velocity (assume the star is made of ideal gas of temperature $\langle T \rangle$).

- Demonstrate that if a binary loses instantaneously an amount of mass greater than half the total mass of the binary, $\Delta M \geq (M_1+M_2)/2$, then the orbit is unbound. This can happen when a supernova goes off in a binary (cf. Blaauw 1961)!

- determine a condition for the minimum mass of a gas cloud to collapse as a function of its temperature and density (**Hint**: collapse $\Leftrightarrow d^2 I/dt^2 < 0$)

Using the model of the $1M_\odot$ star computed earlier with `MESA-web`:

- use one profile file to plot $\Gamma_1 \equiv \Gamma_1(m)$ for a Sun-like star of roughly the same age as the Sun. Label the age of the star for the corresponding profile.

The Sun with no energy sources

Let's use `MESA-web` to revisit the late 19^{th} century/early 20^{th} century debate on the age of the Earth/Sun. (Astro)physicists calculated the Kelvin-Helmholtz timescale and assumed that this was the age of the Sun (from Eq. 19). Geologists and biologists instead argued for a much longer age.

Let's assume we know the age of the Earth to be 4.5×10^9 years (this is what the geologist and biologists argued!), but let's assume, like physicists had to before knowing nuclear physics, that there is no internal energy source in the Sun. In `MESA-web`, there is a `Burning Modifiers` option where you can disable energy release and chemical evolution. Make a model of a $1M_\odot$ star without energy release and/or chemical evolution, until 4.5×10^9 years, and plot an HR diagram. Plot also the $1M_\odot$ star you ran previously (which should have included the nuclear energy release and chemical evolution).

N.B.: we can make a computer code do whatever we want! *Never* take computer simulations as ground truth, they are *at best* only as good as the input!

Write a short paragraph answering the following:

- Which agrees better with the observation of the real Sun at the Earth's age?

- What is its radius and how does it compare to the measured $R_\odot$? (**hint**: you can plot the lines at constant radius on the HRD, or read the outermost radius from the last `profile*.data` file you saved)

- What is the average density of the model without energy generation? (an order of magnitude estimate is sufficient for the purpose).