Computing the Universe, Dynamics IV: Liouville, Boltzmann, Jeans [DRAFT]

Martin Weinberg

June 24, 1998

1 Liouville's Theorem

Liouville's theorem is a fairly straightforward consequence of Hamiltonian dynamics. There are a number of ways to derive this and I will give you a slightly different discussion than the ones found in Goldstein or Landau & Lifshitz, one which is closer to numerical practice.

Assume for simplicity that $H = H(\mathbf{p},\mathbf{q})$ but independent of time. Consider the resulting $2n$-dimensional phase space and note that Hamilton's equations,

$$\begin{aligned} \dot{\mathbf{p}} &= -\frac{\partial H}{\partial \mathbf{q}}, \\ \dot{\mathbf{q}} &= \frac{\partial H}{\partial \mathbf{q}}, \end{aligned}$$

defines a vector field or a *flow* in that phase space; at each point $\mathbf{p},\mathbf{q}$ we can compute a direction. Of course by construction, the flow in this field is the solution to Hamilton's equations. Define a "flow operator", g^t which takes a phase-space point forward by time t. A mathematician would write this as $g^t : (\mathbf{p}(0),\mathbf{q}(0)) \to (\mathbf{p}(t),\mathbf{q}(t))$.

Liouville's theorem states: the flow defined by Hamilton's equations preserves phase-space volume for any region of phase space Γ. In notation, we have

$$g^t\langle\text{volume of }\Gamma\rangle = \langle\text{volume of }\Gamma\rangle.$$

The proof is amazingly simple. Consider g^t as $t \to 0$. Suppose $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. Then $g^t(\mathbf{x}) = \mathbf{x} + \mathbf{f}(\mathbf{x})t + O(t^2)$. Let $V(t) = \langle\text{volume of }\Gamma\rangle$. Then if $\nabla f \equiv 0$, g^t preserves volume. Explicitly,

$$V(t) = \int_{\Gamma(0)} d\mathbf{x} \left|\frac{\partial g^t \mathbf{x}}{\partial \mathbf{x}}\right|$$

expresses the usual construction from multivariate calculus where the quantity in $|\cdot|$ is the Jacobian of the transformation. Now:

$$\frac{\partial g^t(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{I} + \frac{\partial f}{\partial \mathbf{x}} t + O(t^2)$$

where $\mathbf{I}$ is the identity matrix. Therefore:

$$\begin{aligned} \left|\frac{\partial g^t(\mathbf{x})}{\partial \mathbf{x}}\right| &= 1 + t\sum_i \frac{\partial f_i}{\partial x_i} + O(t^2) \\ &= 1 + t\nabla\cdot\mathbf{f} + O(t^2) \end{aligned}$$

Putting this together, we have

$$V(t) = \int_{\Gamma(0)} d\mathbf{x}\left[1 + t\nabla\cdot\mathbf{f} + O(t^2)\right]$$

and therefore

$$\frac{dV(t)}{dt} = \int_{\Gamma(0)} d\mathbf{x}\nabla\cdot\mathbf{f} + O(t) \to_{t\to 0} \int_{\Gamma(0)} d\mathbf{x}\nabla\cdot\mathbf{f}.$$

Finally, for Hamilton's equations,

$$\mathbf{f} = \begin{pmatrix} -\partial H/\partial \mathbf{q} \\ \partial H/\partial \mathbf{p} \end{pmatrix}$$

and therefore

$$\nabla \cdot \mathbf{f} = \frac{\partial}{\partial \mathbf{p}} \cdot \left(\frac{-\partial H}{\partial \mathbf{q}} \right) + \frac{\partial}{\partial \mathbf{q}} \cdot \left(\frac{\partial H}{\partial \mathbf{p}} \right) = 0$$

and that's the proof. Since the flow is defined by the vector **f**, Liouville's theorem is sometimes states as: Hamiltonian follow is incompressible. There are two immediate corollaries, one of which we touched on earlier:

- Phase trajectories don't cross (see earlier as a consequence of Newton's Principle).
- A boundary in phase space always encloses the same group of initial conditions.

2 Liouville's equation

In studying galaxies for example, we would like to be able to apply Liouville's theorem to an ensemble of n points in phase space. Although this is ungainly to start, it turns out that one can make an interesting approximation in the limit $n \to \infty$ to get a simple governing equation for the ensemble. There isn't time to go into this in detail, but I will sketch the ideas here and quote the result.

Define $f(\mathbf{p},\mathbf{q})$ as the distribution of points in phase space. That is, the number of points in some volume dV is given by $dN = f(\mathbf{p},\mathbf{q})d\mathbf{p}\,d\mathbf{q} = f(\mathbf{p},\mathbf{q})dV$. How does dN evolve with time? We have the machinery for this now and we find:

$$\begin{aligned} g^t dN &= g^t f(\mathbf{p},\mathbf{q},t) g^t dV \\ &= f(\mathbf{p},\mathbf{q},t) + \left[\frac{\partial f}{\partial \mathbf{p}} \cdot \left(-\frac{\partial H}{\partial \mathbf{q}} \right) + \frac{\partial f}{\partial \mathbf{q}} \cdot \left(\frac{\partial H}{\partial \mathbf{p}} \right) \right] t dV \end{aligned}$$

where the latter expression is valid in the limit $t \to 0$. Now the total rate of change of dN with time is

$$\begin{aligned}
\frac{dN}{dt} = \frac{d\left[f(\mathbf{p},\mathbf{q})dV\right]}{dt} &= g^t f(\mathbf{p},\mathbf{q},t) g^t dV \\
&= \frac{\partial f(\mathbf{p},\mathbf{q},t)}{\partial t} + \left[\frac{\partial f}{\partial \mathbf{p}} \cdot \left(-\frac{\partial H}{\partial \mathbf{q}}\right) + \frac{\partial f}{\partial \mathbf{q}} \cdot \left(\frac{\partial H}{\partial \mathbf{p}}\right)\right] dV
\end{aligned}$$

or

$$\frac{df(\mathbf{p},\mathbf{q})}{dt} = \frac{\partial f(\mathbf{p},\mathbf{q},t)}{\partial t} + \frac{\partial f}{\partial \mathbf{q}} \cdot \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial f}{\partial \mathbf{p}} \cdot \frac{\partial H}{\partial \mathbf{q}}.$$

If phase space is never created or destroyed (e.g. as in a system of stars or dark matter halo), we have $df/dt = 0$. This is called *Liouville's equation*. Again, notice that $\mathbf{p}$ and $\mathbf{q}$ are arbitrary canonical coordinates and therefore the quantity

$$\frac{\partial H}{\partial \mathbf{p}} \cdot \frac{\partial f}{\partial \mathbf{q}} - \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial f}{\partial \mathbf{p}}$$

must be coordinate independent and is often denoted $[H, f]$ and called the *Poisson bracket*. Liouville's equation then takes the simple form

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + [H, f].$$

Note that this is true for any function of space, not just a phase-space distribution function. In particular, take any f independent of time; f is then an integral of motion and and Liouville's equation takes the form $[H, f] = 0$. In words, *an integral of motion commutes with the Hamiltonian*.

3 The Boltzmann and Vlasov equations

The lectures so far have emphasized principle. We have considered one particle in a given field, ignoring how that field is generated.

We have considered n particles in their mutual field, again assuming that the gravitational potential has some known form. But as astronomers we must confront specific observations and, for example, want to represent a galaxy without the dependence on individual stars. For a fixed total gravitational mass, one way of doing this is taking the limit $n \to \infty$. Then, rather than looking at individual degrees of freedom, we consider the evolution of a distribution of particles in phase space (non-interacting *dust*, if you will). To derive an analog to equations of motion, we begin with Liouville's equation

$$\frac{\partial f^{[N]}}{\partial t} + \left[H, f^{[N]}\right] = 0$$

where the superscript $[N]$ denotes a distribution for N particles: $f^{[N]} = f(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$ where $\mathbf{z} = (\mathbf{p}, \mathbf{q})$. In the limit $N \to \infty$, we expect that the $\mathbf{z}_j$ are completely uncorrelated. The indices are obviously arbitrary, it is sufficient to derive from Liouville's equation an expression that only involves $f^{[1]}$. One begins by integrating Liouville's equation over all coordinates with indices $(2, 3, \ldots, N)$. After a fair bit of algebraic manipulation and careful consideration of correlations, one arrives at the following:

$$\frac{\partial f^{[1]}}{\partial t} + \frac{\partial H}{\partial \mathbf{p}_1} \cdot \frac{\partial f^{[1]}(\mathbf{z}_1, t)}{\partial \mathbf{q}_1} - \frac{\partial H}{\partial \mathbf{q}_1} \cdot \frac{\partial f^{[1]}(\mathbf{z}_1, t)}{\partial \mathbf{p}_1} = 0.$$

This equation governs the evolution of the single particle distribution function in the large-scale field of the entire system. One often uses the shorthand notation $f = f^{[1]}$ and writes

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + [H, f].$$

This is called the *collisionless Boltzmann equation* (or sometimes the *Vlasov equation*). In this form it looks just like the Liouville's equation but the derivation is not a trivial one. The procedure which I have glossed over here is known as the $1/N$ expansion or the BBGKY hierarchy[1]. If we retain the first term in $1/N$, which involves the correlated distribution of two particles, we have the Fokker-Planck equation which is important in modeling globular clusters, galactic nuclei and other dense systems where the collisionless approximation does not apply.

4 Jeans' Theorem

Although the similarity between the collisionless Boltzmann equation and the Liouville equation is deep, the coincidence in form allows us to exploit all of the same mathematical machinery. The most important of these consequences is called Jeans' Theorem which says:

1. Any steady-state solution of the collisionless Boltzmann equation depends on phase space through the integrals of motion of orbits in the gravitational potential.
2. Any function of the integrals of motion satisfy the time-independent collisionless Boltzmann equation.

The proof is both short and illustrative:

1. If f is a steady-state solution, f is an integral of the motion and therefore only depends itself on integrals of the motion. Let depend on the independent variables y_i: $f = f(\{y_i\})$. Using the Liouville form, we have
$$\frac{df}{dt} = 0 = \sum_i \frac{\partial f}{\partial y_i}\frac{dy_i}{dt} = \sum_i \frac{\partial f}{\partial y_i}[H, y_i].$$

[1] I recall that Bogoliubov, Born, Green, Kirkwood & Yvon all worked on this at about the same time.

Since the y_i are independent, this can only be true if $[H, y_i] = 0$; in other words, the y_i are constants of the motion.

2. Assume that $f = f(I_1, \ldots, I_5)$ (we can have at most 5 constants of the motion since time-independence implies energy conservation). Then,

$$\frac{df}{dt} = \sum_{i=1}^{5} \frac{\partial f}{\partial I_i} \frac{dI_i}{dt}$$

and therefore $df/dt = 0$ since all the I_i are constants of the motion.

As an important practical consequence, then, if we know $f(\mathbf{I})$ for a given system, the problem is completely solved! Unfortunately, the theorem does not tell us what these integrals are or how many of them there are. Fortunately, we already have some insight. In systems where action-angle variables describe all orbits in phase space (e.g. we can solve the H-J equation for the gravitationally bound system), the actions are constants of the motion and the angles vary linearly with time. In fact, one can prove that a time-independent distribution function must depend on actions alone; and, in general, the number of constants of the motion necessary to describe the system will be equal to its dimension. There are some notable exceptions that occur when the frequencies are degenerate. In this case, we may transform to a new set of variables where the one or more of the angle oscillation frequencies are zero. Then the angles are conserved too. The most notable case is Kepler's problem which has three degenerate frequencies and therefore two *extra* constants of the motion. These correspond the orbit being closed and non-precessing.

5 Astronomical examples and associated distribution functions

5.1 One-dimensional slab

Distribution of stars varies only in the vertical direction. Therefore, the distribution function $f = f(E) = \tilde{f}(I_z)$ since $E = H(I_z)$.

5.2 Two-degree of freedom problem: motion in a disk

Here there are two degrees of freedom. We have seen discussed in the context of orbit classification the interpretation of the radial and azimuthal actions. The distribution function is then $f = f(I_r, I_\phi) = \tilde{f}(E, h)$.

5.3 Three-degree of freedom problem: motion in a sphere (e.g. globular cluster or giant cluster galaxy)

This is quite similar to the disk problem if the velocity distribution has no preferred axis. Because there are two tangential directions, the third action is another angular-momentum, I_θ and the distribution function is often written $f(I_r, I_\phi, I_\theta) = \tilde{f}(E, J, J_z)$, where J is the total angular momentum of the orbit and J_z is its projection along the z-axis.

6 Jeans' equations