

11. Introduction to Liouville's Theorem

Michael Fowler

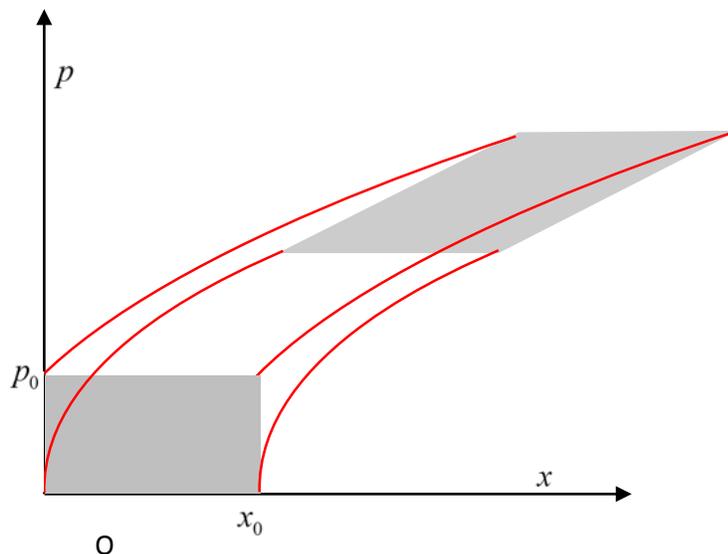
Paths in Simple Phase Spaces: the SHO and Falling Bodies

Let's first think further about paths in phase space. For example, the simple harmonic oscillator, with Hamiltonian $H = p^2 / 2m + m\omega^2 q^2 / 2$, describes circles in phase space parameterized with the variables $(p, m\omega q)$. (A more usual notation is to write the potential term as $\frac{1}{2}kq^2$.)

Question: are these circles the only possible paths for the oscillator to follow?

Answer: yes: any other path would intersect a circle, and at that point, with both position and velocity defined, there is only one path forward (and back) in time possible, so the intersection can't happen.

Here's an example from Taylor of paths in phase space: four identical falling bodies are released



Four phase space trajectories for falling objects

simultaneously, see figure, x measures distance vertically down. Two are released with zero momentum from the origin O and from a point x_0 meters down, the other two are released with initial momenta p_0 , again from the points O, x_0 . (Note the difference in initial slope.)

Check: convince yourself that all these paths are parts of parabolas centered on the x -axis. (Just as the simple harmonic oscillator phase space is filled with circular paths, this one is filled with parabolas.)

Bodies released with the same initial velocity at the same time will keep the same vertical distance apart, those released with different initial velocities will keep the same velocity difference, since all accelerate at g . Therefore, *the area of the parallelogram formed by the four phase space points at a later time will have the same area as the initial square.*

Exercise: convince yourself that all the points of an initial vertical side of the square all stay in line as time goes on, even though the line does not stay vertical.

The four sides of the square deform with time to the four sides of the parallelogram, point by point. This means that if we have a falling stone corresponding initially to a point inside the square, it will go to a point inside the parallelogram, because if somehow its path reached the boundary, we would have two

paths in phase space intersecting, and a particle at one point in phase space has a uniquely defined future path (and past).

Following Many Systems: a “Gas” in Phase Space

We’ve looked at four paths in phase space, corresponding to four falling bodies, all beginning at $t = 0$, but with different initial co-ordinates in (p, x) . Suppose now we have many falling bodies, so that at $t = 0$ a region of phase space can be imagined as filled with a “gas” of points, each representing one falling body, initially at (p_i, x_i) , $i = 1, \dots, N$.

The argument above about the phase space path of a point within the square at $t = 0$ staying inside the square as time goes on and the square distorts to a parallelogram must also be true for any dynamical system, and any closed volume in phase space, since it depends on phase space paths never intersecting: that is,

if at $t = 0$ some closed surface in phase space contains a number of points of the gas, those same points remain inside the surface as it develops in time -- none exit or enter.

For the number of points N sufficiently large, the phase space time development looks like the flow of a fluid.

Liouville’s Theorem: Local Gas Density Is Constant along a Phase Space Path

The falling bodies phase space square has one more lesson for us: visualize now a uniformly dense gas of points inside the initial square. Not only does the gas stay within the distorting square, the area it covers in phase space remains constant, as discussed above, so *the local gas density stays constant as the gas flows through phase space.*

Liouville’s theorem is that this constancy of local density is true for general dynamical systems.

Landau’s Proof Using the Jacobian

Landau gives a very elegant proof of elemental volume invariance under a general canonical transformation, proving the Jacobian multiplicative factor is always unity, by clever use of the generating function of the canonical transformation.

Jacobians have wide applicability in different areas of physics, so this is a good time to review their basic properties, which we do below, as a preliminary to giving the proof.

It must be admitted that there are simpler ways of deriving Liouville’s theorem, directly from Hamilton’s equations, *the reader may prefer to skip the Jacobian proof at first reading.*

Jacobian for Time Evolution

As we’ve established, time development is equivalent to a canonical coordinate transformation,

$$(p_t, q_t) \rightarrow (p_{t+\tau}, q_{t+\tau}) \equiv (P, Q).$$

Since we already know that the number of points inside a closed volume is constant in time, Liouville's theorem is proved if we can show that the volume enclosed by the closed surface is constant, that is, with V' denoting the volume V evolves to become, we must prove

$$\int_{V'} dQ_1 \dots dQ_s dP_1 \dots dP_s = \int_V dq_1 \dots dq_s dp_1 \dots dp_s ?$$

If you're familiar with Jacobians, you know that (by definition)

$$\int dQ_1 \dots dQ_s dP_1 \dots dP_s = \int D dq_1 \dots dq_s dp_1 \dots dp_s$$

where the Jacobian

$$D = \frac{\partial(Q_1, \dots, Q_s, P_1, \dots, P_s)}{\partial(q_1, \dots, q_s, p_1, \dots, p_s)}.$$

Liouville's theorem is therefore proved if we can establish that $D = 1$. If you're not familiar with Jacobians, or need reminding, read the next section!

Jacobians 101

You can skip this section if you're already familiar with Jacobians.

Suppose we are integrating a function over some region of ordinary three-dimensional space,

$$I = \int_V f(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

but we want to change variables of integration to a different set of coordinates (q_1, q_2, q_3) such as, for example, (r, θ, ϕ) . The new coordinates are of course functions of the original ones $q_i(x_1, x_2, x_3)$, etc., and we assume that in the region of integration they are smooth, well-behaved functions. We can't simply re-express f in terms of the new variables, and replace the volume differential $dx_1 dx_2 dx_3$ by $dq_1 dq_2 dq_3$, that gives the wrong answer—in a plane, you can't replace $dx dy$ with $dr d\theta$, you have to use $r dr d\theta$. That extra factor r is called the Jacobian, it's clear that in the plane a small element with sides of fixed lengths $(\delta r, \delta \theta)$ is bigger the further it is from the origin, not all $\delta r \delta \theta$ elements are equal, so to speak. Our task is to construct the Jacobian for a general change of coordinates.

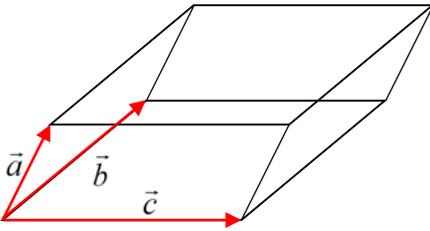
We need to think carefully about the volumes in the three-dimensional space represented by $dx_1 dx_2 dx_3$ and by $dq_1 dq_2 dq_3$. Of course, the x_i 's are just ordinary perpendicular Cartesian axes so the volume is just the product of the three sides of the little box, $dx_1 dx_2 dx_3$. Imagine this little box, its corner closest to the origin at (x_1, x_2, x_3) and its furthest point at the other end of the body diagonal at $(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3)$. Let's take these two points in the q_i coordinates to be at (q_1, q_2, q_3) and

$(q_1 + dq_1, q_2 + dq_2, q_3 + dq_3)$. In visualizing this, bear in mind that the q axes need not be perpendicular to each other (but they cannot all lie in a plane, that would not be well-behaved).

For the x coordinate integration, we imagine filling the space with little cubical boxes. For the q integration, we have a system of space filling infinitesimal parallelepipeds, in general pointing different ways in different regions (think (r, θ)). What we need to find is the volume of the incremental parallelepiped with sides we'll write as vectors in x -coordinates, $d\vec{q}_1, d\vec{q}_2, d\vec{q}_3$. These three incremental vectors are along the corresponding q coordinate axes, and the three added together are the displacement from (x_1, x_2, x_3) to $(x_1 + dx_1, x_2 + dx_2, x_3 + dx_3) \equiv (q_1 + dq_1, q_2 + dq_2, q_3 + dq_3)$.

Hence, in components,

$$d\vec{q}_1 = \left(\frac{\partial q_1}{\partial x_1} dx_1, \frac{\partial q_1}{\partial x_2} dx_2, \frac{\partial q_1}{\partial x_3} dx_3 \right).$$



Now the volume of the parallelepiped with sides the three vectors from the origin $\vec{a}, \vec{b}, \vec{c}$ is $\vec{a} \cdot \vec{b} \times \vec{c}$ (recall $|\vec{b} \times \vec{c}|$ is the area of the parallelogram, then the dot product singles out the component of \vec{a} perpendicular to the plane of \vec{b}, \vec{c}).

So, the volume corresponding to the increments dq_1, dq_2, dq_3 in q space is

$$d\vec{q}_1 \cdot d\vec{q}_2 \times d\vec{q}_3 = \begin{vmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_1}{\partial x_2} & \frac{\partial q_1}{\partial x_3} \\ \frac{\partial q_2}{\partial x_1} & \frac{\partial q_2}{\partial x_2} & \frac{\partial q_2}{\partial x_3} \\ \frac{\partial q_3}{\partial x_1} & \frac{\partial q_3}{\partial x_2} & \frac{\partial q_3}{\partial x_3} \end{vmatrix} dx_1 dx_2 dx_3 = D dx_1 dx_2 dx_3,$$

writing D (Landau's notation) for the **determinant**, which is in fact the **Jacobian**, often denoted by J .

The standard notation for this determinantal Jacobian is

$$D = \frac{\partial(q_1, q_2, q_3)}{\partial(x_1, x_2, x_3)},$$

So the appropriate replacement for the three dimensional incremental volume element represented in the integral by $dq_1 dq_2 dq_3$ is

$$dq_1 dq_2 dq_3 \rightarrow \frac{\partial(q_1, q_2, q_3)}{\partial(x_1, x_2, x_3)} dx_1 dx_2 dx_3.$$

The **inverse**

$$D^{-1} = \frac{\partial(x_1, x_2, x_3)}{\partial(q_1, q_2, q_3)},$$

this is easily established using the chain rule for differentiation.

Exercise: check this!

Thus the change of variables in an integral is accomplished by rewriting the integrand in the new variables, and replacing

$$I = \int_V f(x_1, x_2, x_3) dx_1 dx_2 dx_3 = \int_V f(q_1, q_2, q_3) \frac{\partial(x_1, x_2, x_3)}{\partial(q_1, q_2, q_3)} dq_1 dq_2 dq_3.$$

The argument in higher dimensions is just the same: on going to dimension $n + 1$, the hypervolume element is equal to that of the n dimensional element multiplied by the component of the new vector perpendicular to the n dimensional element. The determinantal form does this automatically, since a determinant with two identical rows is zero, so in adding a new vector only the component perpendicular to all the earlier vectors contributes.

We've seen that the chain rule for differentiation gives the inverse as just the Jacobian with numerator and denominator reversed, it also readily yields

$$\frac{\partial(x_1, x_2, x_3)}{\partial(q_1, q_2, q_3)} \cdot \frac{\partial(q_1, q_2, q_3)}{\partial(r_1, r_2, r_3)} = \frac{\partial(x_1, x_2, x_3)}{\partial(r_1, r_2, r_3)},$$

and this extends trivially to n dimensions.

It's also evident from the determinantal form of the Jacobian that

$$\frac{\partial(x_1, x_2, x_3)}{\partial(q_1, q_2, x_3)} = \frac{\partial(x_1, x_2)}{\partial(q_1, q_2)},$$

identical variables in numerator and denominator can be canceled. Again, this extends easily to n dimensions.

Jacobian proof of Liouville's Theorem

After this rather long detour into Jacobian theory, recall we are trying to establish that the volume of a region in phase space is unaffected by a canonical transformation, we need to prove that

$$\int dQ_1 \dots dQ_s dP_1 \dots dP_s = \int dq_1 \dots dq_s dp_1 \dots dp_s,$$

and that means we need to show that the Jacobian

$$D = \frac{\partial(Q_1, \dots, Q_s, P_1, \dots, P_s)}{\partial(q_1, \dots, q_s, p_1, \dots, p_s)} = 1.$$

Using the theorems above about the inverse of a Jacobian and the chain rule product,

$$D = \frac{\partial(Q_1, \dots, Q_s, P_1, \dots, P_s)}{\partial(q_1, \dots, q_s, p_1, \dots, p_s)} / \frac{\partial(q_1, \dots, q_s, p_1, \dots, p_s)}{\partial(Q_1, \dots, Q_s, P_1, \dots, P_s)}.$$

Now invoking the rule that if the same variables appear in both numerator and denominator, they can be cancelled,

$$D = \left\{ \frac{\partial(Q_1, \dots, Q_s)}{\partial(q_1, \dots, q_s)} \right\}_{P=\text{constant}} / \left\{ \frac{\partial(p_1, \dots, p_s)}{\partial(P_1, \dots, P_s)} \right\}_{q=\text{constant}}.$$

Up to this point, the equations are valid for any nonsingular transformation—but to prove the numerator and denominator are equal in this expression requires that the equation be canonical, that is, be given by a generating function, as explained earlier.

Recall now the properties of the generating function $\Phi(q, P, t)$,

$$d\Phi(q, P, t) = d\left(F + \sum P_i Q_i\right) = \sum p_i dq_i + \sum Q_i dP_i + (H' - H)dt,$$

from which

$$p_i = \partial\Phi(q, P, t) / \partial q_i, \quad Q_i = \partial\Phi(q, P, t) / \partial P_i, \quad H' = H + \partial\Phi(q, P, t) / \partial t.$$

In the expression for the Jacobian D , the i, k element of the numerator is $\partial Q_i / \partial q_k$.

In terms of the generating function $\Phi(q, P)$ this element is $\partial^2\Phi / \partial q_k \partial P_i$.

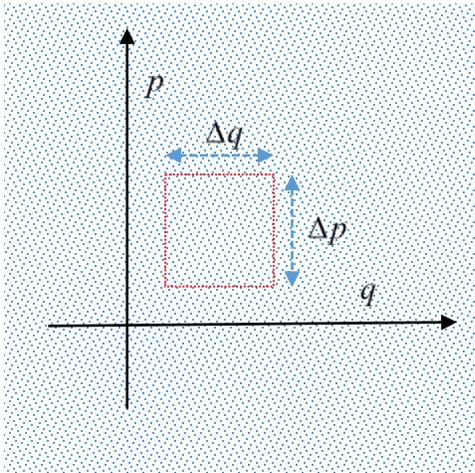
Exactly the same procedure for the denominator gives the i, k element to be $\partial P_i / \partial p_k = \partial^2\Phi / \partial q_i \partial P_k$.

In other words, the two determinants are the same (rows and columns are switched, but that doesn't affect the value of a determinant). This means $D = 1$, and Liouville's theorem is proved.

Simpler Proof of Liouville's Theorem

Landau's proof given above is extremely elegant: since phase space paths cannot intersect, point inside a volume stay inside, no matter how the volume contorts, and since time development is a canonical transformation, the total volume, given by integrating over volume elements $dqdp$, stays the same, since it's an integral over the corresponding volume elements $dQdP$ and we've just shown that $dQdP = dqdp$.

Here we'll take a slightly different point of view: we'll look at a small square in phase space and track how its edges are moving, to prove its volume isn't changing. (We'll stick to one dimension, but the generalization is straightforward.)



The points here represent a "gas" of many systems in the two dimensional (q, p) phase space, and with a small square area $\Delta q, \Delta p$, tagged by having all the systems on its boundary represented by dots of a different color. What is the incremental change in area of this initially square piece of phase space in time dt ?

Begin with the top edge: the particles are all moving with velocities (\dot{q}, \dot{p}) , but of course the only change in *area* comes from the \dot{p} term, that's the outward movement of

the boundary, so the area change in dt from the movement of this boundary will be $\dot{p}\Delta q dt$.

Meanwhile, there will be a similar term from the bottom edge, and the *net* contribution, top plus bottom edges, will depend on the change in \dot{p} from bottom to top, that is, a net area change from movement of these edges $(\partial\dot{p} / \partial p)\Delta p\Delta q dt$.

Adding in the other two edges (the sides), with an exactly similar argument, the total area change is

$$(\partial\dot{p} / \partial p + \partial\dot{q} / \partial q)\Delta p\Delta q dt .$$

But from Hamilton's equations $\dot{p} = \partial H / \partial q$, $\dot{q} = -\partial H / \partial p$, so

$$\partial\dot{p} / \partial p = \partial^2 H / \partial p \partial q, \quad \partial\dot{q} / \partial q = -\partial^2 H / \partial p \partial q$$

and therefore

$$\partial\dot{p} / \partial p + \partial\dot{q} / \partial q = 0,$$

establishing that the total incremental area change as the square distorts is zero.

The conclusion is that the flow of the gas of systems in phase space is like an incompressible fluid, *but* with one important qualification: the density may vary with position! It just doesn't vary along a dynamical path.

Energy Gradient and Phase Space Velocity

For a time-independent Hamiltonian, the path in phase space (q, p) is a constant energy line, and we can think of the whole phase space as delineated by many such lines, exactly analogous to contour lines joining points at the same level on a map of uneven terrain, energy corresponding to height above sea level. The gradient at any point, the vector pointing exactly uphill and therefore perpendicular to the constant energy path, is

$$\vec{\nabla} H = (\partial H / \partial q, \partial H / \partial p),$$

here $H = E$. The velocity of a system's point moving through phase space is

$$\vec{v} = (\dot{q}, \dot{p}) = (\partial H / \partial p, -\partial H / \partial q).$$

This vector is perpendicular to the gradient vector, as it must be, of course, since the system moves along a constant energy path. But, interestingly, it has the same magnitude as the gradient vector!

What is the significance of that? Imagine a small square sandwiched between two phase space paths close together in energy, and suppose the distance between the two paths is decreasing, so the square is getting squeezed, at a rate equal to the rate of change of the energy gradient. But at the same time it must be getting stretched along the direction of the path, an amount equal to the rate of change of phase space velocity along the path—and they are equal. So, this is just Liouville again, its area doesn't change.

