

Why is momentum the generator of translations?

Santiago Quintero de los Ríos
for homotopico.com

September 23, 2018

In quantum mechanics, it is often stated that *the momentum [operator] is the generator of translations*. However, most texts fail to give a satisfactory justification to this claim, if any is given at all. In some cases, the proof is simply begging the question. In others, the burden of proof is shifted to classical mechanics, stating that “this was already proved in classical mechanics”. Again, most standard texts of classical mechanics fail to give a clear proof of this claim, although the mathematical apparatus is there. What we will do here is define *precisely* what it means for observable to be a generator of a group of transformations, both in quantum and classical mechanics, and prove that, indeed, the momentum observable *is* the generator of translations in classical mechanics.

The origin of the question

We define the translation operator \hat{T}_a acting on elements $\psi \in L^2(\mathbb{R})$ (the Hilbert space of square-integrable complex functions over \mathbb{R}) by, of course, translating the wavefunction a distance a to the right:

$$(\hat{T}_a \psi)(x) := \psi(x - a). \quad (1)$$

This operator is actually a whole family of operators that depend on the parameter a . It is clear that \hat{T}_0 is the identity operator. Now since this family of operators is particularly well-behaved¹, it has an **infinitesimal generator**, which is some operator \hat{w} such that

$$\hat{T}_a = \exp(-ia\hat{w}). \quad (2)$$

Yes, I know that the standard name for this generator is \hat{p} , but I am trying to remove any possible meaning to this operator. For the time being, we only know that \hat{w} is the generator of translations. That’s why I don’t call it \hat{p} .

What does \hat{w} look like? One way to calculate the infinitesimal generator of a family of operators is by evaluating the derivative at $a = 0$. If we expand the right-hand side of equation 2 as a power series, then

$$\exp(-ia\hat{w}) = \sum_{n=0}^{\infty} \frac{1}{n!} (-ia\hat{w})^n, \quad (3)$$

then it follows that

$$\left. \frac{d}{da} \right|_{a=0} \hat{T}_a = \left. \frac{d}{da} \right|_{a=0} \exp(-ia\hat{w}) = -i\hat{w}. \quad (4)$$

So let ψ be any function on $L^2(\mathbb{R})$. We have that

$$\left. \frac{d}{da} \right|_{a=0} \hat{T}_a \psi(x) = \left. \frac{d}{da} \right|_{a=0} \psi(x - a) = -\frac{d}{dx} \psi(x), \quad (5)$$

¹That is, it is a one-parameter absolutely continuous group.

so we can say that the infinitesimal generator is

$$\hat{w} = -i \frac{d}{dx}. \quad (6)$$

And hey, look at that. It turns out that this is what we call the momentum operator in quantum mechanics.

However, it is not clear **at all** why the operator $-i \frac{d}{dx}$ should be given the highly suggestive name of “momentum”. So this it’s not that the generator of translations is momentum because it has the form $-i \frac{d}{dx}$, but the other way around:

We say that $-i \frac{d}{dx}$ is the momentum operator, because it is the one that generates translations.

Therefore, the question is shifted: **Why do we give the name “momentum” to the operator that generates translations?**

The usual answers

If you only want a half-plausible argument, here it goes. Let \hat{x} be the position operator, that is the expectation value $\langle \psi | \hat{x} | \psi \rangle$ is the expected value for the position of a particle in state ψ . It can be shown that the position operator acts on wavefunctions by multiplication by the variable x , i.e.

$$(\hat{x}\psi)(x) = x\psi(x). \quad (7)$$

This equation can also be used to define \hat{x} . Now we want to compute the commutator $[\hat{x}, \hat{w}]$. On one hand, we have:

$$(\hat{x}\hat{w}\psi)(x) = -ix \frac{d}{dx} \psi(x), \quad (8)$$

but on the other hand,

$$(\hat{w}\hat{x}\psi)(x) = -i \frac{d}{dx} (x\psi(x)) = -i \left(\psi(x) + x \frac{d}{dx} \psi(x) \right). \quad (9)$$

Then, clearly

$$[\hat{x}, \hat{w}] \psi(x) = i\psi(x), \quad (10)$$

and so $[\hat{x}, \hat{w}] = i\hat{I}$. Now we consider the “quantization rule” that turns *classical* observables f into Hermitian operators \hat{f} and that turns Poisson brackets *almost* into Lie brackets

$$\{f, g\} \mapsto i [\hat{f}, \hat{g}], \quad (11)$$

and what we do is go backwards! We found an operator \hat{w} that satisfies the commutation relation $[\hat{x}, \hat{w}] = i\hat{I}$, so the *classical* observables they come from satisfy the Poisson-bracket relation $\{x, w\} = 1$. We know that \hat{x} is the position operator, so x must be the position observable. On the other hand, we know that for the canonical momentum p ,

$$\{x, p\} = \frac{\partial x}{\partial x} \frac{\partial p}{\partial p} - \frac{\partial x}{\partial p} \frac{\partial p}{\partial x} = 1. \quad (12)$$

We can clearly see that $\{x, p\} = \{x, w\} = 1$, so this suggests that $w = p$! And we’re done. w is the momentum, so the operator \hat{w} is the momentum operator.

But wait a minute...

The relation

$$\{x, w\} = 1 \quad (13)$$

is not enough to *define* w ! This is because the operator $f \mapsto \{x, f\}$ is degenerate. As a trivial example,

$$\{x, p + x\} = 1, \quad (14)$$

and actually, for any function $f(x)$ that *does not depend on* p , it follows that

$$\{x, p + f(x)\} = 1. \quad (15)$$

This means that w is an observable of the form $w = p + f(x)$ for *some* function $f(x)$. At this point you might think “eh, close enough”. And fair enough, then you’re done!

But I’m not thoroughly (actually, not at all) satisfied. There is another clue though. This previous argument is one of two that I’ve seen in QM books. The other one is as follows:

In classical mechanics it is shown that momentum is the generator of translations.

And well, yes, that could be enough for our purposes. We can’t expect every new theory to work out of the box and be able to stand by itself, right? When we construct a new theory from a mathematical perspective, we look back to other theories that are well-established and we understand, look for analogies and similarities, and interpret the symbols in such a way that everything is as consistent as possible with what came before.

Therefore it is quite reasonable to say that this operator \hat{w} is the momentum operator if we already know from classical mechanics that the generator of translations is the momentum. We’ve shown that in quantum mechanics, translations *also* have infinitesimal generators, so it is natural to interpret those as momenta.

Again, that is **if** we already know from classical mechanics that momentum is the generator of translations. And I think I missed that lecture? That’s a big “if”.

That quote up there shifts the burden of proof to classical mechanics. So what we’re going to do is plunge into classical mechanics and try to show that momentum is the generator of translations from two different points of view: the “standard” Goldstein point-of-view, and the more modern symplectic geometry point-of-view. We will see from both that we can reasonably show that, in a certain way, momentum is the generator of translations.

The standard point of view

More precisely, what we will prove is the following:

To each (classical) observable f we can *canonically* assign a one-parameter group of *canonical* transformations Φ_a that preserve f . In this case, we say that f generates Φ_a . When we choose the canonical momentum $f = p$, the resulting transformations are translations, and thus we say that p is the generator of translations.

In this first section, we will do it in a bit of a dirty way. If infinitesimals make you uncomfortable (completely understandable), then always note that when we write, e.g. $\tilde{x} = x + \epsilon X$, what we mean is that $\tilde{x} = \tilde{x}(\epsilon)$ is a function of ϵ , that $\tilde{x}(0) = x$, and $X = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \tilde{x}(\epsilon)$. If you would like to see a more concise proof that requires a bit of differential topology, skip to the next section.

Consider some system with phase space \mathcal{S} , with canonical coordinates q^i, p_j , and let f be an arbitrary observable. We want to find a *canonical* transformation² $Q^i = Q^i(q, p)$, $P_j = P_j(q, p)$ that preserves f and depends on only one parameter ϵ . For small ϵ , we can write

$$\begin{aligned} Q^i(\epsilon) &= q^i + \epsilon A^i \\ P_j(\epsilon) &= p_j + \epsilon B_j \end{aligned} \quad (16)$$

²And an additional question that I would like to explore is “what’s the huge deal about canonical transformations”? We know that they preserve the structure of the Hamilton-Jacobi equations, but not necessarily the Hamiltonian, so that doesn’t seem like much, does it? Well it turns out that the *structure* itself of the equations has an interesting and beautiful mathematical background... But I’m saving that for later.

with A^i, B_j functions of q, p that we want to determine. Since we want these transformations to be canonical, the Poisson brackets must be conserved, so

$$\begin{aligned}\{q^i, p_i\} &= 1 = \{Q^i, P_i\} \\ &= \{q^i + \epsilon A^i, p_i + \epsilon B_i\} \\ &= \{q^i, p_i\} + \epsilon \left(\{A^i, p_i\} + \{q^i, B_i\} \right) + \mathcal{O}(\epsilon^2).\end{aligned}\tag{17}$$

Then, working up to first order in ϵ , we require that

$$\{A^i, p_i\} + \{q^i, B_i\} = 0.\tag{18}$$

Unravel these Poisson brackets,

$$\{A^i, p_i\} = \sum_k \left(\frac{\partial A^i}{\partial q^k} \frac{\partial p_i}{\partial p_k} - \frac{\partial A^i}{\partial p_k} \frac{\partial p_i}{\partial q^k} \right) = \frac{\partial A^i}{\partial q^i}\tag{19}$$

$$\{q^i, B_i\} = \sum_k \left(\frac{\partial q^i}{\partial q^k} \frac{\partial B_i}{\partial p_k} - \frac{\partial q^i}{\partial p_k} \frac{\partial B_i}{\partial q^k} \right) = \frac{\partial B_i}{\partial p_i},\tag{20}$$

then we require A^i, B_i to be such that

$$\frac{\partial A^i}{\partial q^i} + \frac{\partial B_i}{\partial p_i} = 0.\tag{21}$$

One way to guarantee that condition (21) is satisfied is by finding some *other* observable, say g , and write

$$\begin{aligned}A^i &= \frac{\partial g}{\partial p_i} \\ B_j &= -\frac{\partial g}{\partial q^j}.\end{aligned}\tag{22}$$

However, not every g will work, since we also require the transformation to preserve the observable f . So again, for small ϵ , we can write

$$\begin{aligned}f(Q^i, P_j) &= f(q^i + \epsilon A^i, p_j + \epsilon B_j) \\ &= f(q^i, p_j) + \epsilon \sum_k \left(A^k \frac{\partial f}{\partial q^k} + B_k \frac{\partial f}{\partial p_k} \right) + \mathcal{O}(\epsilon^2).\end{aligned}\tag{23}$$

Since we require the transformation to preserve f , then $f(Q^i, P_j) = f(q^i, p_j)$ and thus

$$\sum_k \left(A^k \frac{\partial f}{\partial q^k} + B_k \frac{\partial f}{\partial p_k} \right) = 0.\tag{24}$$

But again, if we assume that A^i, B_j come from an observable g as in equation (22), this condition becomes

$$\sum_k \left(\frac{\partial g}{\partial p^k} \frac{\partial f}{\partial q^k} - \frac{\partial g}{\partial q^k} \frac{\partial f}{\partial p_k} \right) = \{f, g\} = 0.\tag{25}$$

Therefore the admissible observables g that can be used for equation (22) are such that $\{f, g\} = 0$. Which one is the *absolute simplest one* to choose? Well f itself! This works since $\{f, f\} = 0$. Therefore the *natural* choice is $g = f$, so that the transformation *for infinitesimal* ϵ is

$$\begin{aligned} Q^i(\epsilon) &= q^i + \epsilon \frac{\partial f}{\partial p^i} \\ P_j(\epsilon) &= p_j - \epsilon \frac{\partial f}{\partial q^j} . \end{aligned} \tag{26}$$

This equation only holds for infinitesimally small ϵ , since we've been working only up to first order. What we want now is to write the exact solution to large ϵ . To do so, note that equation (26) gives us a *differential equation* for Q and P : writing $q^i = Q^i(0)$, $p_j = P_j(0)$, it can be easily seen that

$$\begin{aligned} \frac{dQ^i}{d\epsilon} &= \frac{\partial f}{\partial p^i} \\ \frac{dP_j}{d\epsilon} &= -\frac{\partial f}{\partial q^j} . \end{aligned} \tag{27}$$

Note that these are just like the Hamilton-Jacobi equations, except that “time” is ϵ and the “Hamiltonian” is f !

Now we see how any general observable changes when we do an infinitesimal transformation (26). Let g be any observable. Then, if we start at a point (q_0, p_0) , and write $(q_1, p_1) = (Q(\epsilon), P(\epsilon))$ (I'm dropping the i, j indices), then

$$\begin{aligned} g(Q^i(\epsilon), P_j(\epsilon)) &= g\left(q_0^i + \epsilon \left.\frac{\partial f}{\partial p_j}\right|_{(q_0, p_0)}, p_{0,j} - \epsilon \left.\frac{\partial f}{\partial q^j}\right|_{(q_0, p_0)}\right) \\ g(q_1, p_1) &= g(q_0, p_0) + \epsilon \sum_k \left(\left.\frac{\partial f}{\partial p_k}\right|_{(q_0, p_0)} \left.\frac{\partial g}{\partial q^k}\right|_{(q_0, p_0)} - \left.\frac{\partial f}{\partial q^k}\right|_{(q_0, p_0)} \left.\frac{\partial g}{\partial p_k}\right|_{(q_0, p_0)} \right) + \mathcal{O}(\epsilon^2) \\ g(q_1, p_1) &= g(q_0, p_0) + \epsilon \{g, f\}|_{(q_0, p_0)} + \mathcal{O}(\epsilon^2) . \end{aligned} \tag{28}$$

How do we go to large ϵ ? Well, the last equation tells us that

$$\frac{d}{d\epsilon} g(Q(\epsilon), P(\epsilon)) = \{g, f\}_{Q(\epsilon), P(\epsilon)} , \tag{29}$$

therefore, if we write $g(\epsilon) = g(Q(\epsilon), P(\epsilon))$, taking the derivative with respect to ϵ again we obtain:

$$\frac{d^2}{d\epsilon^2} g(\epsilon) = \frac{d}{d\epsilon} \{g, f\}_\epsilon = \{\{g, f\}, f\}_\epsilon . \tag{30}$$

So if we call X_f the differential operator $X_f(g) := \{g, f\}$, we can see that

$$\frac{d^n}{d\epsilon^n} g(\epsilon) = X_f(X_f(\dots X_f(g)))_\epsilon = X_f^n(g)_\epsilon . \tag{31}$$

And thus, by Taylor-expanding g with respect to ϵ around $\epsilon = 0$, we see that

$$g(\epsilon) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n}{d\epsilon^n} \right|_{\epsilon=0} g(\epsilon) = \sum_{n=0}^{\infty} \frac{1}{n!} X_f^n(g)_{(q_0, p_0)} = \exp(\epsilon X_f(0)) g . \tag{32}$$

We call X_f the **infinitesimal generator** associated to f .

Alright. We should stop here for a bit and gather what we have. We have shown that for any observable f , we can find a family of canonical transformations, which we write as $Q(\epsilon), P(\epsilon)$, given

by the set of Hamilton-Jacobi-like equations (27). Note that if we fix a starting point (q_0, p_0) , this family of transformations gives rise to a *curve* in phase space, given by $q = Q(\epsilon)$, $p = P(\epsilon)$. Any other observable g also changes along this curve as $g(\epsilon) = \exp(\epsilon X_f(0))g$, or rather as $\frac{dg}{d\epsilon} = \{g, f\}$. In particular, f is constant along this curve, since $\frac{df}{d\epsilon} = \{f, f\} = 0$. Therefore the family of transformations is canonical and preserves f , as was desired.

Now here comes what was promised. If we choose $f = p_j$, the momentum observable, then the infinitesimal generator associated to p_j , X_{p_j} , is the differential operator

$$X_{p_j}(g) = \{g, p_j\} = \sum_k \frac{\partial g}{\partial q^k} \frac{\partial p_j}{\partial p_k} - \frac{\partial g}{\partial p_k} \frac{\partial p_j}{\partial q^k} = \frac{\partial}{\partial q^j}(g). \quad (33)$$

Ring any bells?

Furthermore, the family of transformations is the solution to the Hamilton-Jacobi-like equations (27) with $f = p_j$, i.e.

$$\begin{aligned} \frac{dQ^i}{d\epsilon} &= \frac{\partial p_j}{\partial p^i} = \delta_j^i \\ \frac{dP_k}{d\epsilon} &= -\frac{\partial p_j}{\partial q^k} = 0, \end{aligned} \quad (34)$$

which is readily integrated, given an initial point $q_0 = Q(0)$, $p_0 = P(0)$:

$$Q^j(\epsilon) = q_0^j + \epsilon; \quad (35)$$

and all the other coordinates constant. This is indeed a translation in the j -th space coordinate.

So there you have it! In classical mechanics, to each observable f we assign a differential operator X_f that generates one-parameter canonical transformations that preserve f . In the case where $f = p$, the differential operator is $X_p = \frac{\partial}{\partial q}$, and the canonical transformations are translations! Therefore we say that **momentum is the generator of translations**.

The mathematician's way

This way is basically exactly the same as the one before, just with the more sophisticated language of differential topology, specifically in the context of symplectic manifolds. It can be easily seen that the results given here are exactly the same as the results given in the previous section, when we write them in Darboux coordinates.

So first, some definitions.

A **symplectic manifold** is a smooth manifold \mathcal{M} of dimension $2n$ endowed with a closed, non-degenerate 2-form ω , which we call a **symplectic form**. It can be shown that for each point $x \in \mathcal{M}$ there exists a neighborhood U of x and coordinates $q^1, \dots, q^n, p_1, \dots, p_n$ such that the symplectic form can be written as

$$\omega = dq^\mu \wedge dp_\mu. \quad (36)$$

Here we are using Einstein's sum notation. This result is **Darboux's theorem**, and the coordinates (q^μ, p_μ) are called **Darboux coordinates**.

Since ω is non-degenerate, it induces an isomorphism $(-)_b : T_p \mathcal{M} \rightarrow T_p^* \mathcal{M}$, given pointwise by the relation

$$v_b(u) := \omega(v, u) \quad (37)$$

for all $u, v \in T_p \mathcal{M}$ and all $p \in \mathcal{M}$. The inverse of $(-)_b$ is denoted by $(-)^{\sharp} : T_p^* \mathcal{M} \rightarrow T_p \mathcal{M}$.

This is basically all we need! Let $f \in C^\infty(\mathcal{M})$. Then we can define the **Hamiltonian vector field associated to f** , denoted by $X_f \in \mathfrak{X}(\mathcal{M})$, as

$$X_f = (df)^{\sharp}. \quad (38)$$

By this definition, for any other vector field Y , we have that

$$\omega(X_f, Y) = df(Y) = Y[f]. \quad (39)$$

Therefore an alternative definition for X_f is the unique vector field such that $\iota_{X_f}\omega = df$.

Using Cartan's magic formula, we see that

$$L_{X_f}\omega = \iota_{X_f}(d\omega) + d(\iota_{X_f}\omega) = 0. \quad (40)$$

Here, the first term is zero since ω is closed (i.e. $d\omega = 0$), and the second one is zero too since $\iota_{X_f}\omega = df$, and $d^2 = 0$. In addition to this, the directional derivative of f along X_f is zero, since

$$X_f(f) = df(X_f) = \iota_{X_f}\omega(X_f) = \omega(X_f, X_f) = 0. \quad (41)$$

This means that if Φ_τ^f is the flow of X_f , then for all values of τ

$$\left(\Phi_\tau^f\right)^* \omega = \omega, \quad (42)$$

and $f \circ \Phi_\tau^f = f$. That is, neither f nor ω change along the integral curves of the Hamiltonian vector field associated to f .

Furthermore, since $\Phi_\tau^f : \mathcal{M} \rightarrow \mathcal{M}$ is a diffeomorphism, then it satisfies the conditions for being a **symplectomorphism** or **canonical transformation**. Namely, a diffeomorphism $f : \mathcal{M} \rightarrow \mathcal{N}$ between two symplectic manifolds (\mathcal{M}, ω) and (\mathcal{N}, Ω) is a symplectomorphism if $f^*\Omega = \omega$.

What we have now is the following: to any $f \in C^\infty(\mathcal{M})$ we can assign a Hamiltonian vector field $X_f \in \mathfrak{X}(\mathcal{M})$ whose flow Φ_τ^f is a symplectomorphism that preserves f . Now if we work locally in Darboux coordinates (q^μ, p_μ) , so that $\omega = dq^\mu \wedge dp_\mu$. If we write X_f locally as

$$X_f = A^\mu \frac{\partial}{\partial q^\mu} + B_\nu \frac{\partial}{\partial p_\nu}, \quad (43)$$

then we have that

$$\begin{aligned} \iota_{X_f}\omega &= \iota_{X_f}(dq^\mu \wedge dp_\mu) \\ &= dq^\mu(X_f)dp_\mu - dq^\mu dp_\mu(X_f) \\ &= A^\mu dp_\mu - B_\mu dq^\mu = df. \end{aligned} \quad (44)$$

However, since

$$df = \frac{\partial f}{\partial q^\mu} dq^\mu + \frac{\partial f}{\partial p_\mu} dp_\mu,$$

then necessarily the components of X_f must be

$$\begin{aligned} A^\mu &= \frac{\partial f}{\partial p_\mu} \\ B_\mu &= -\frac{\partial f}{\partial q^\mu}. \end{aligned} \quad (45)$$

The Hamiltonian vector field associated to f is then, in Darboux coordinates,

$$X_f = \frac{\partial f}{\partial p_\mu} \frac{\partial}{\partial q^\mu} - \frac{\partial f}{\partial q^\mu} \frac{\partial}{\partial p_\mu} = \{-, f\}. \quad (46)$$

And once again, if we choose $p_\nu = f$, then the Hamiltonian vector field is

$$X_{p_\nu} = \frac{\partial}{\partial q^\nu}, \quad (47)$$

So that the flow of X_{p_ν} is simply translation along the q^ν coordinate. Once again, we can say that the [Hamiltonian vector field associated to the] canonical momentum is the generator of translations.

The takeaway

We proved a general result in classical mechanics: to any observable f , we can assign a differential operator $X_f = \{-, f\}$, called the infinitesimal generator or Hamiltonian vector field associated to f , that *generates* canonical transformations that preserve f . The one-parameter group of canonical transformations associated to X_f is given by $\Phi_\tau^f = \exp(\tau X_f)$, and we say that f is the generator of Φ^f .

In particular, if we choose the canonical momentum p to be the observable f , then the infinitesimal generator is $X_p = \frac{d}{dq}$ and the group of canonical transformations associated to X_p is precisely the group of translations of the q coordinate. Therefore, we can say that **the momentum is the generator of translations**.

Now in quantum mechanics, we simply **define** the momentum operator \hat{p} to be the infinitesimal generator of the unitary group of translations, and we see that this definition is consistent with canonical quantization.

References

- Sakurai, J. J. (1995). *Modern Quantum Mechanics, Revised Edition*.
- Goldstein, H. , Poole, C. & Safko, J. (2001). *Classical Mechanics, Third Edition*.
- José, J. V. & Saletan, E. (1998). *Classical Dynamics: A Contemporary Approach*. This book is great. I love it. Read it.
- Abraham, R. & Marsden, J. (2008). *Foundations of Mechanics*.

Great thanks to Laura Arboleda for checking style and consistency.