Predicting the future with Newton's Second Law

To represent the motion of an object (ignoring rotations for now), we need three functions $x(t)$, $y(t)$, and $z(t)$, which describe the spatial coordinates of the object for each possible time. Since there are an infinite number of possible times (between any starting and ending time), giving the coordinates at all possible times is actually an infinite amount of information!

Part of the magic of physics is that we can predict all of this information knowing only the initial position $\vec{r}(t=0) = (x(t=0), y(t=0), z(t=0))$ and velocity $\vec{v}(t=0) = (v_x(t=0), v_y(t=0), v_z(t=0))$ of an object, plus the information about the object's environment (specifically, what forces are acting on it). The key tool is Newton's Second Law, which we write as:[1]

$$\vec{a} = \frac{1}{m}\vec{F}_{NET} \; . \tag{1}$$

This says that *we can predict the acceleration of an object by knowing the forces on the object.* We are assuming that the object's environment is understood well enough that we can predict the forces on the object from the object's position and velocity.

Why we can predict the future

To see why Newton's Second Law allows us to predict the future, we need to remember that acceleration is defined to be the rate of change of velocity $\vec{a} = d\vec{v}/dt$. So given the acceleration $\vec{a}$ at some time t, we can say that in a time δt, the velocity will change by $\vec{a}\delta t$. If the velocity at time t is $\vec{v}(t)$, the velocity at the later time $(t+\delta t)$ will then be

$$\vec{v}(t+\delta) \approx \vec{v}(t) + \delta t \; \vec{a} \; . \tag{2}$$

In exactly the same way, using the definition of velocity as the rate of change of position, we get

$$\vec{r}(t+\delta) \approx \vec{r}(t) + \delta t \; \vec{v} \; . \tag{3}$$

We put approximately equals to ($\approx$) here because the acceleration and/or velocity might be changing with time. In this case, the equations become exact only in the limit where $\delta t \to 0$, but in practice, we just need to take δt small enough to get whatever accuracy we desire.

Let's understand why (2) and (3) together with Newton's Law (1) are so powerful. Remembering that Newton's law allows us to predict acceleration from the force on the object, and that the force at some time is determined by the position and velocity of the object at that time (we'll write the force as $\vec{F}(\vec{r}(t), \vec{v}(t))$ to remind ourselves of this), we can summarize (2) and (3) as

$$\vec{v}(t+\delta) \;\; \approx \;\; \vec{v}(t) + \delta t \, \frac{1}{m}\vec{F}(\vec{r}(t), \vec{v}(t))$$

[1] Here, we are assuming that speed is small compared with the speed of light, so that we can use $\vec{p} = m\vec{v}$ to rewrite Newton's Law from its original form $d\vec{p}/dt = \vec{F}$ to the familiar $\vec{F} = m\vec{a}$.

$$\vec{r}(t+\delta) \approx \vec{r}(t) + \delta t\, \vec{v} \,.$$

Here, the right hand sides are all things that we can calculate given the position and velocity of the object at time t. The left hand sides are the position at velocity at a slightly later time. So if we know position and velocity now, we can predict what they will be slightly later! We can do this as many times as we want to predict what the object will do in the future.

To summarize, we get the following recipe for predicting the motion.

- Start with the position $\vec{r}(t)$ and the velocity $\vec{v}(t)$ at some time.
- Given these, determine the net force on the object.
- Using (4), calculate the position and velocity $\vec{r}(t+\delta t)$ and the velocity $\vec{v}(t+\delta t)$ at some slightly later time.
- Repeat.

By taking the time step δt to be sufficiently small, we can predict the position and velocity of an object at some specified later time with arbitrary precision.[2] In many cases, there are simpler ways to actually predict the motion than using this repetitive procedure. But our discussion here shows that we can always predict the motion *in principle* no matter how complicated the forces are.

Differential equations of motion

The equations (4) are equivalent to the following exact equations[3]

$$\begin{aligned} \frac{d\vec{v}}{dt} &= \frac{1}{m}\vec{F}(\vec{r}(t), \vec{v}(t)) \\ \frac{d\vec{r}}{dt} &= \vec{v} \,. \end{aligned}$$

Since these equations contain derivatives, they are known as DIFFERENTIAL EQUATIONS. The solutions to equations like this are functions (in this case $\vec{r}(t)$ and $\vec{v}(t)$). What we have seen is that given some INITIAL CONDITIONS – that is, the position and velocity $\vec{r}(t_1)$ and $\vec{v}(t_1)$ at some initial time t_1 – there will be a unique solution (that we can find in principle using the repetitive method) to these equations. This solution describes the motion of the object at all later times. For this reason, the equations (4) are known as the EQUATIONS OF MOTION for the object.[4]

[2] Note that in some systems, for large enough times into the future, achieving such precision would require taking a δt that is impractically small.

[3] Here the second one is just the definition of velocity and the first combines the definition of acceleration with Newton's first Law. These were exactly what we used to derive (4)

[4] Sometimes, the two equations (4) are combined into one by using the second to eliminate v from the first. This gives

$$\frac{d^2\vec{r}}{dt^2} = \frac{1}{m}\vec{F}(\vec{r}(t), \vec{v}(t))$$

, but the equations of motion are a little less clear to interpret in this form.

Methods for solving the equations of motion

Let's now discuss various procedures for actually solving the equations of motion.

1) The Euler method

In the most difficult cases (actually, in most realistic cases), the repetitive procedure we have described in the previous section (or some closely related procedure) may be the only way of predicting the motion. While this procedure is tedious to do by hand, it is very easy to implement on a computer. The specific procedure described is known as the Euler method for solving the differential equations. If we predict the position and/or velocity of an object at some later time using this method, the accuracy of our result (difference from the exact result) is generally proportional to the size of our time step δt. For some other numerical methods, the error decreases more quickly as we take δt to zero.

2) Solving the differential equations directly

In some cases, the differential equations are simple enough that we can solve them directly by finding a family of functions that satisfies the equations and then choosing the ones that have the right initial conditions. Various higher-level math courses (or books on differential equations) teach you techniques for doing this. In the simplest cases, it's possible just to guess a solution and check that it works. As an example, if we had $dv/dt = -Bv$ (e.g. for a resistive force proportional to the speed of the object), we could recall that exponential functions have a derivative proportional to the function itself and then guess the family of solutions $v(t) = Ae^{-Bt}$, using the initial velocity to determine the constant A.

3) Cases where the force is some known function of time - antidifferentiation method

A special case where we can directly find a solution is when the force is some known function of time, which *does not explicitly refer to the position or the velocity of the object.*[5] This includes cases where the forces are constant, but also cases where they can change with time. In these cases, the equations of motion simplify to

$$\begin{aligned} \frac{dv_x}{dt} &= A_x(t) \\ \frac{dx}{dt} &= v_x \; . \end{aligned}$$

where $A_x(t)$ is known, and we may have similar equations for y and z. Note that $A_x(t)$ *is not allowed to depend on $\vec{v}$ or $\vec{r}$.* In this case we can use the following method:

[5]Examples of forces that do depend explicitly on position or velocity are the drag force (with magnitude proportional to v^2), or the force from a spring, which changes as the object moves and the spring stretches.

- Find a function $g(t)$ whose derivative is $A_x(t)$.
- Since the derivative of $v_x(t)$ is also $A_x(t)$, the graphs of $g(t)$ and $v_x(t)$ have the same slope everywhere. This means that $v_x(t) = g(t) + C$ for some constant C.
- To find C, we use the information about what v_x is at the initial time. For example, if we know v_x is v_1 at time t_1, then we need to choose $C = v(t_1) - g(t_1)$, so that
$$v_x(t) = g(t) - g(t_1) + v(t_1) \; . \tag{4}$$
- Once we know $v_x(t)$ we repeat the same procedure to find $x(t)$.
- Use the same approach independently for y and z if necessary.

4) Cases where the force is some known function of time - area method

Remarkably, there is another equivalent method for determining the velocity at later times if we are given the acceleration as a function of time. This may be summarized as
$$v_x(t_2) = v_x(t_1) + \int_{t_1}^{t_2} A_x(t)dt \; , \tag{5}$$
where the expression $\int_{t_1}^{t_2} A_x(t)dt$ means the *area under the the graph of $A_x(t)$ between t_1 and t_2*. Similarly, once we know $v_x(t)$, we can write
$$x(t_2) = x(t_1) + \int_{t_1}^{t_2} v_x(t)dt \; .$$
In practice, we might use this method if the acceleration is given to us in the form of a graph of a very simple function for which we can immediately read off the area.

Origin of the area method

Where does this come from? It's actually just the Euler method in disguise! Let's say we have
$$\frac{dv}{dt} = A(t) \; , \tag{6}$$
and the velocity at time t_1 is equal to v_1. Then we can use (2) to write the velocity at time $v(t_1 + \delta t)$ as
$$v(t_1 + \delta t) = v(t_1) + \delta t \; A(t_1) \; .$$
Repeating for the later time $t_1 + 2\delta t$, we get
$$\begin{aligned} v(t_1 + 2\delta t) &= v(t_1 + \delta t) + \delta t \; A(t_1 + \delta t) \\ &= v(t_1) + \delta t \; A(t_1) + \delta t \; A(t_1 + \delta t) \end{aligned}$$

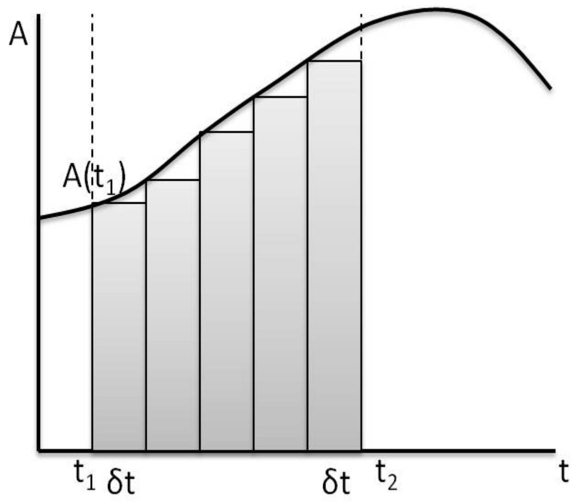

Figure 1: Rectangles of thickness δt under graph of $A(t)$ from t_1 to t_2. Area of the first rectangle is $\delta t\ A(t_1)$, area of the second rectangle is $\delta t\ A(t_1 + \delta t)$, and so forth.

where we have used our previous result in the last line. If we keep repeating this all the way to some later time t_2, we'll get

$$v(t_2) = v(t_1) + \delta t A(t_1) + \delta t\ A(t_1 + \delta t) + \delta t\ A(t_1 + 2\delta t) + \ldots + \delta t\ A(t_2 - \delta t)\ .$$

Now, looking at figure 1, we notice that each term here with a δt is exactly the area of one of the rectangles under the graph. In the limit where δt goes to zero (where the Euler method becomes exact), the rectangles become infinitely thin and completely fill in the region under the curve between t_1 and t_2 (i.e. there are no gaps). Thus, the sum of the areas is the area under the curve, and we have derived the formula (5).

The fundamental theorem of calculus

We've actually just figured out a completely AMAZING MATHEMATICAL FACT. Combining what we've learned so far, we've actually come up with a method for *how to calculate the area under the curve of a function*! Let's review: we saw in the previous section that if v satisfies the equation (6), then the change in v from time t_1 to time t_2 is equal to the area under the graph of $A(t)$ from t_1 to t_2. But we'd already given a completely different method for finding this change in v. According to the other method, if we find a function $g(t)$ whose derivative is $A(t)$, then the change in velocity from t_1 to t_2 is $g(t_2) - g(t_1)$ (using equation (4). Since both methods are correct, it must be that

$$\int_{t_1}^{t_2} A(t)dt = g(t_2) - g(t_1)$$

where g is any function whose derivative is A(t). This amazing result connects differentiation with integration (finding areas) and is part of THE FUNDAMENTAL THEOREM OF CALCULUS.

Example: constant acceleration

The methods above all work when the acceleration is constant, and reproduce all the standard kinematics formulae for constant acceleration. To see this, suppose we have an object with initial position x_0 and velocity v_0 at time $t = 0$, and suppose that the acceleration is a constant value A. Then we have

$$\frac{dv}{dt} = A$$

so using the antidifferentiation method, we have

$$v(t) = At + C \; .$$

where C is a constant. To get the correct value $v = v_0$ at $t = 0$, we must have $C = v_0$, so we get

$$v(t) = v_0 + At \; .$$

From this, we have

$$\frac{dx}{dt} = v_0 + At \; ,$$

so using antidifferentiation again, we find

$$x(t) = C' + v_0 t + \frac{1}{2}At^2 \; .$$

Since $x(0) = x_0$, we must have $C' = x_0$, so we finally obtain the familiar formula

$$x(t) = x_0 + v_0 t + \frac{1}{2}At^2 \; .$$

It is important to note that this (and other formulae like $v_1^2 = v_0^2 + 2Ad$) are *only correct if acceleration is constant*!