

The Principle of Least Action

A special lecture—almost verbatim (addition to Chapter 19 Volume III of Feynman Lectures)

When I was in high school, my physics teacher—whose name was Mr. Bader —called me down one day after physics class and said, ‘You look bored; I want to tell you something interesting.’ Then he told me something which I found absolutely fascinating, and have, since then, always found fascinating. Every time the subject comes up, I work on it. In fact, when I began to prepare this lecture I found myself making more analyses on the thing. Instead of worrying about the lecture, I got involved in a new problem. The subject is this—the principle of least action.

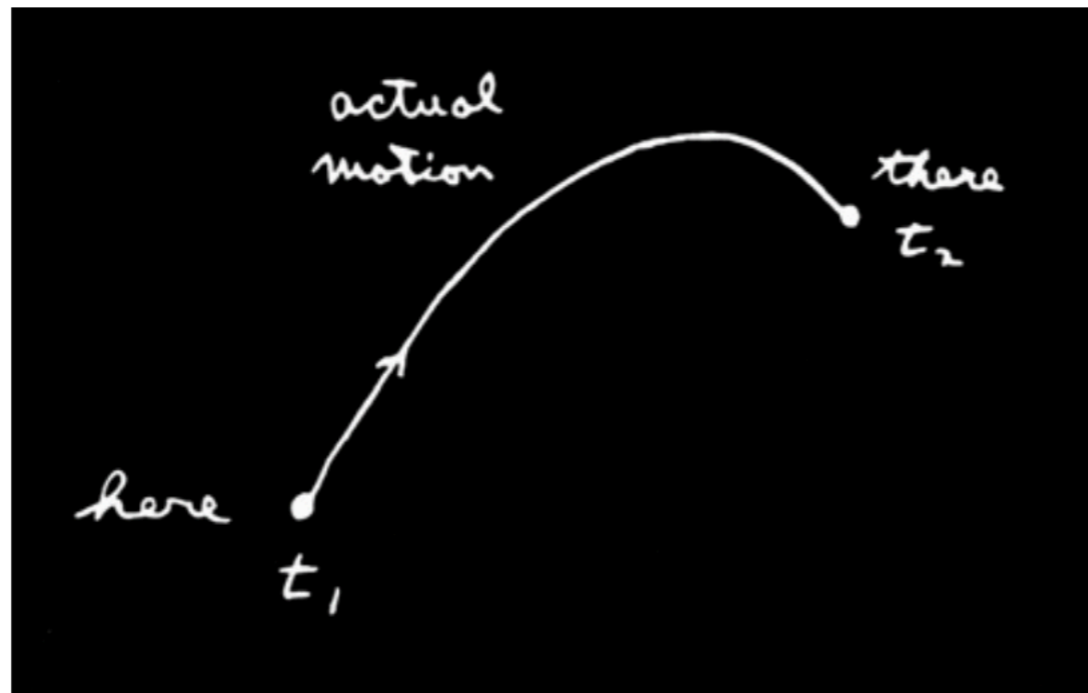


Fig. 19-1.

Mr. Bader told me the following: Suppose you have a particle (in a gravitational field, for instance) which starts somewhere and moves to some other point by free motion—you throw it, and it goes up and comes down (Fig. 19-1). It goes from the original place to the final place in a certain amount of time. Now, you try a different motion. Suppose that to get from here to there, it went as shown in Fig. 19-2 but got there in just the same amount of time.

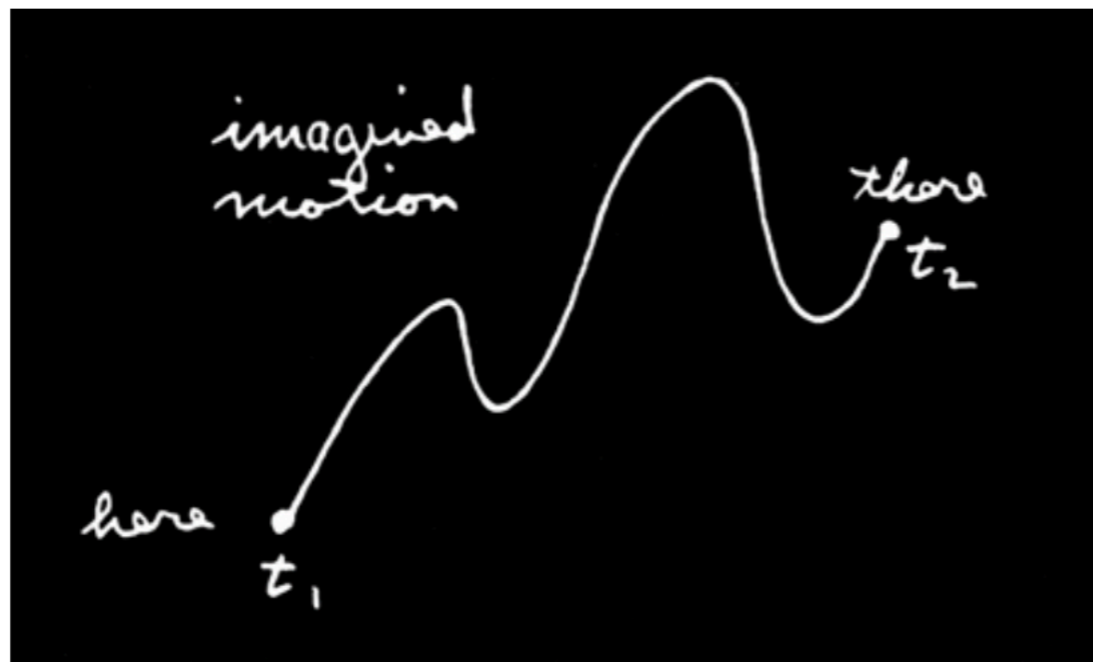


Fig. 19-2.

Then he said this: If you calculate the kinetic energy at every moment on the path, take away the potential energy, and integrate it over the time during the whole path, you'll find that the number you'll get is *bigger* than that for the actual motion.

In other words, the laws of Newton could be stated not in the form $F = ma$ but in the form: the average kinetic energy less the average potential energy is as little as possible for the path of an object going from one point to another.

Let me illustrate a little bit better what it means. If you take the case of the gravitational field, then if the particle has the path $x(t)$ (let's just take one dimension for a moment; we take a trajectory that goes up and down and not sideways), where x is the height above the ground, the kinetic energy is $1/2 m(dx/dt)^2$, and the potential energy at any time is mgx . Now I take the kinetic energy minus the potential energy at every moment along the path and integrate that with respect to time from the initial time to the final time. Let's suppose that at the original time we started at some height and at the end of the time we are definitely ending at some other place (Fig. 19-3).

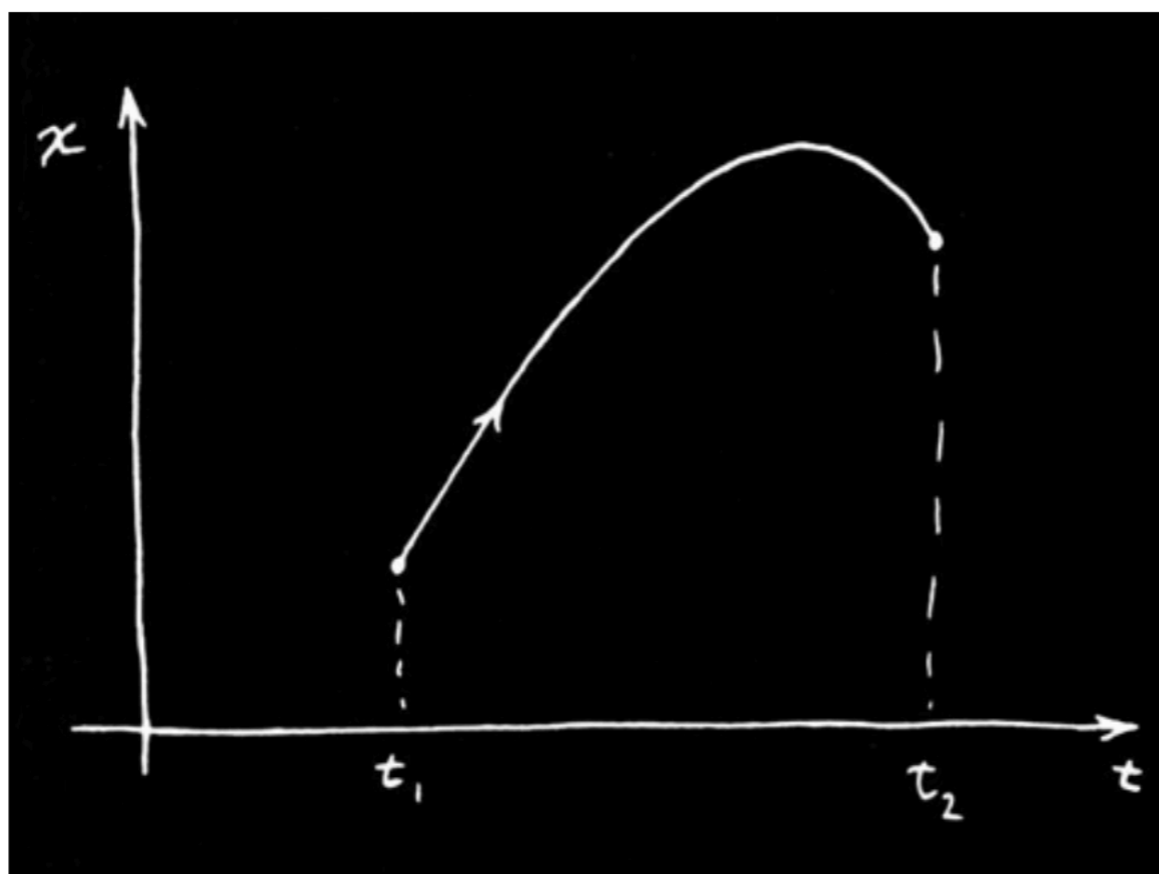


Fig. 19-3.

Then the integral is

$$\int_{t_1}^{t_2} \left[\frac{1}{2} m \left(\frac{dx}{dt} \right)^2 - mgx \right] dt.$$

The actual motion is some kind of a curve—it's a parabola if we plot against the time—and gives a certain value for the integral. But we could *imagine* some other motion that went very high and came up and down in some peculiar way (Fig. 19-4).

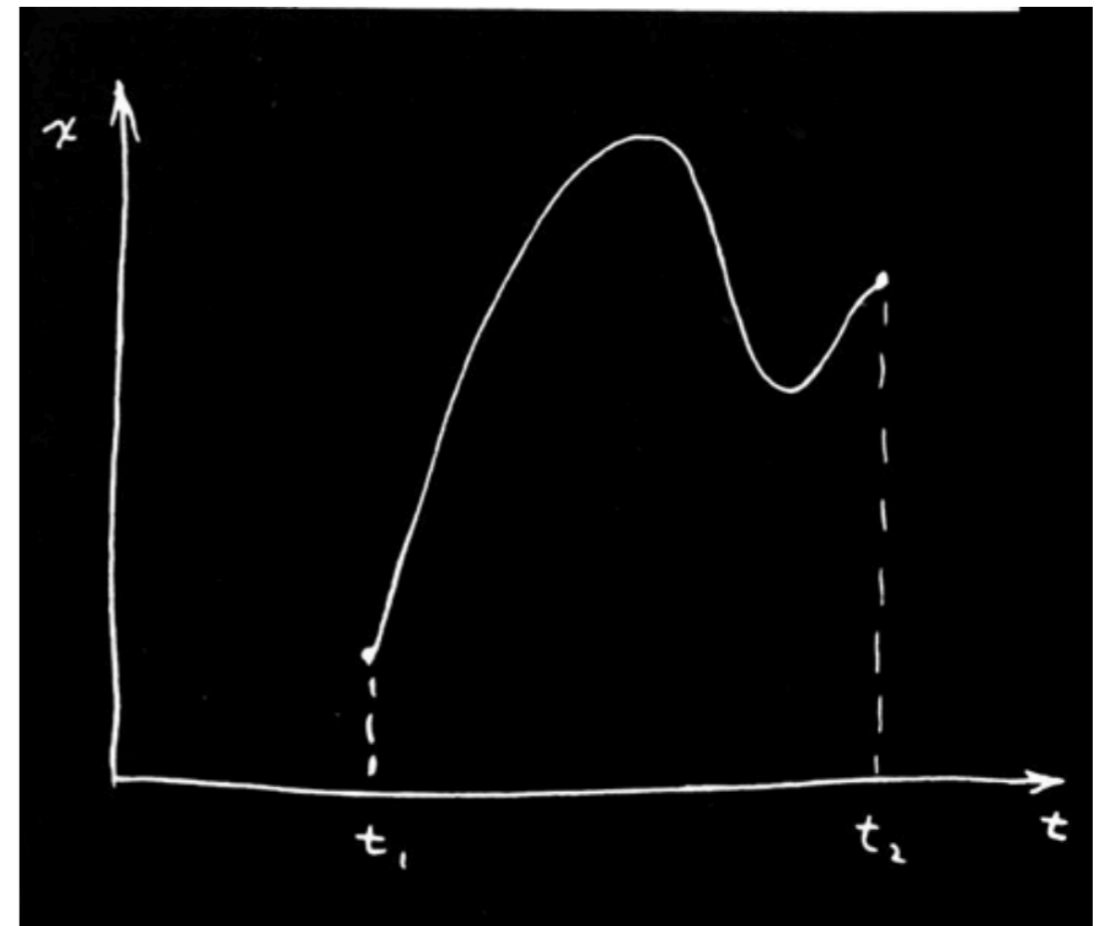


Fig. 19-4.

We can calculate the kinetic energy minus the potential energy and integrate for such a path ... or for any other path we want. The miracle is that the true path is the one for which that integral is least.

Let's try it out. First, suppose we take the case of a free particle for which there is no potential energy at all. Then the rule says that in going from one point to another in a given amount of time, the kinetic energy integral is least, so it must go at a uniform speed. (We know that's the right answer—to go at a uniform speed.) Why is that? Because if the particle were to go any other way, the velocities would be sometimes higher and sometimes lower than the average. The average velocity is the same for every case because it has to get from 'here' to 'there' in a given amount of time.

As an example, say your job is to start from home and get to school in a given length of time with the car. You can do it several ways: You can accelerate like mad at the beginning and slow down with the brakes near the end, or you can go at a uniform speed, or you can go backwards for a while and then go forward, and so on. The thing is that the average speed has got to be, of course, the total distance that you have gone over the time. But if you do anything but go at a uniform speed, then sometimes you are going too fast and sometimes you are going too slow. Now the mean square of something that deviates around an average, as you know, is always greater than the square of the mean; so the kinetic energy integral would always be higher if you wobbled your velocity than if you went at a uniform velocity. So we see that the integral is a minimum if the velocity is a constant (when there are no forces). The correct path is shown in Fig. 19-5.

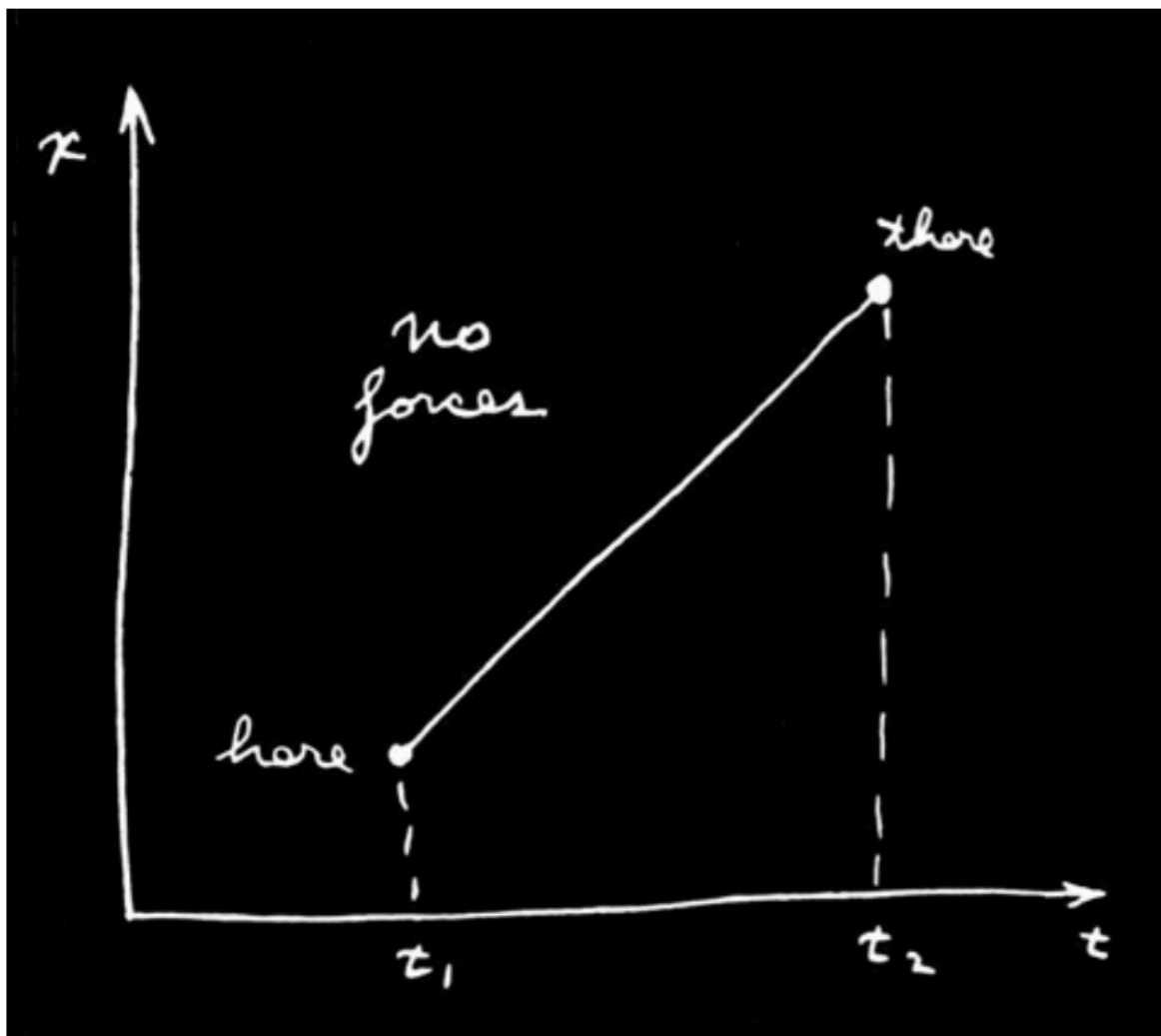


Fig. 19-5.

Now, an object thrown up in a gravitational field does rise faster first and then slow down. That is because there is also the potential energy, and we must have the least difference of kinetic and potential energy on the average. Because the potential energy rises as we go up in space, we will get a lower difference if we can get as soon as possible up to where there is a high potential energy. Then we can take that potential away from the kinetic energy and get a lower average. So it is better to take a path which goes up and gets a lot of negative stuff from the potential energy (Fig. 19-6).

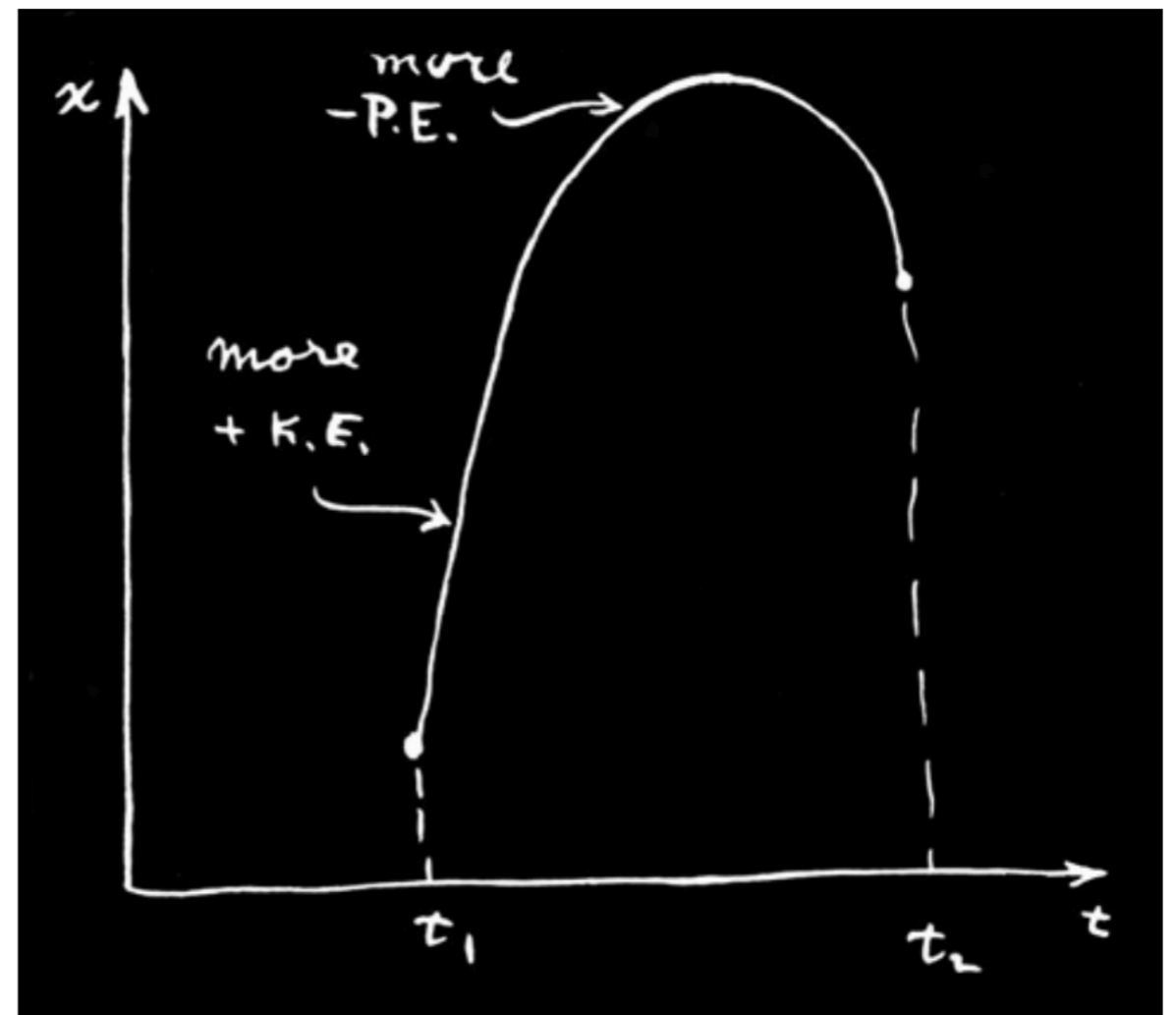


Fig. 19-6.

On the other hand, you can't go up too fast, or too far, because you will then have too much kinetic energy involved—you have to go very fast to get way up and come down again in the fixed amount of time available. So you don't want to go too far up, but you want to go up some. So it turns out that the solution is some kind of balance between trying to get more potential energy with the least amount of extra kinetic energy—trying to get the difference, kinetic minus the potential, as small as possible.

That is all my teacher told me, because he was a very good teacher and knew when to stop talking. But I don't know when to stop talking. So instead of leaving it as an interesting remark, I am going to horrify and disgust you with the complexities of life by proving that it is so. The kind of mathematical problem we will have is very difficult and a new kind. We have a certain quantity which is called the *action*, S . It is the kinetic energy, minus the potential energy, integrated over time.

$$\text{Action} = S = \int_{t_1}^{t_2} (\text{KE} - \text{PE}) dt.$$

Remember that the PE and KE are both functions of time. For each different possible path you get a different number for this action. Our mathematical problem is to find out for what curve that number is the least.

You say—Oh, that's just the ordinary calculus of maxima and minima. You calculate the action and just differentiate to find the minimum.

But watch out. Ordinarily we just have a function of some variable, and we have to find the value of that *variable* where the function is least or most. For instance, we have a rod which has been heated in the middle and the heat is spread around.

For each point on the rod we have a temperature, and we must find the point at which that temperature is largest. But now for each path in space we have a number—quite a different thing—and we have to find the path in space for which the number is the minimum. That is a completely different branch of mathematics. It is not the ordinary calculus. In fact, it is called the *calculus of variations*.

There are many problems in this kind of mathematics. For example, the circle is usually defined as the locus of all points at a constant distance from a fixed point, but another way of defining a circle is this: a circle is that curve of given length which encloses the biggest area. Any other curve encloses less area for a given perimeter than the circle does. So if we give the problem: find that curve which encloses the greatest area for a given perimeter, we would have a problem of the calculus of variations—a different kind of calculus than you're used to.

So we make the calculation for the path of an object. Here is the way we are going to do it. The idea is that we imagine that there is a true path and that any other curve we draw is a false path, so that if we calculate the action for the false path we will get a value that is bigger than if we calculate the action for the true path (Fig. 19-7).

Problem: Find the true path. Where is it? One way, of course, is to calculate the action for millions and millions of paths and look at which one is lowest. When you find the lowest one, that's the true path.

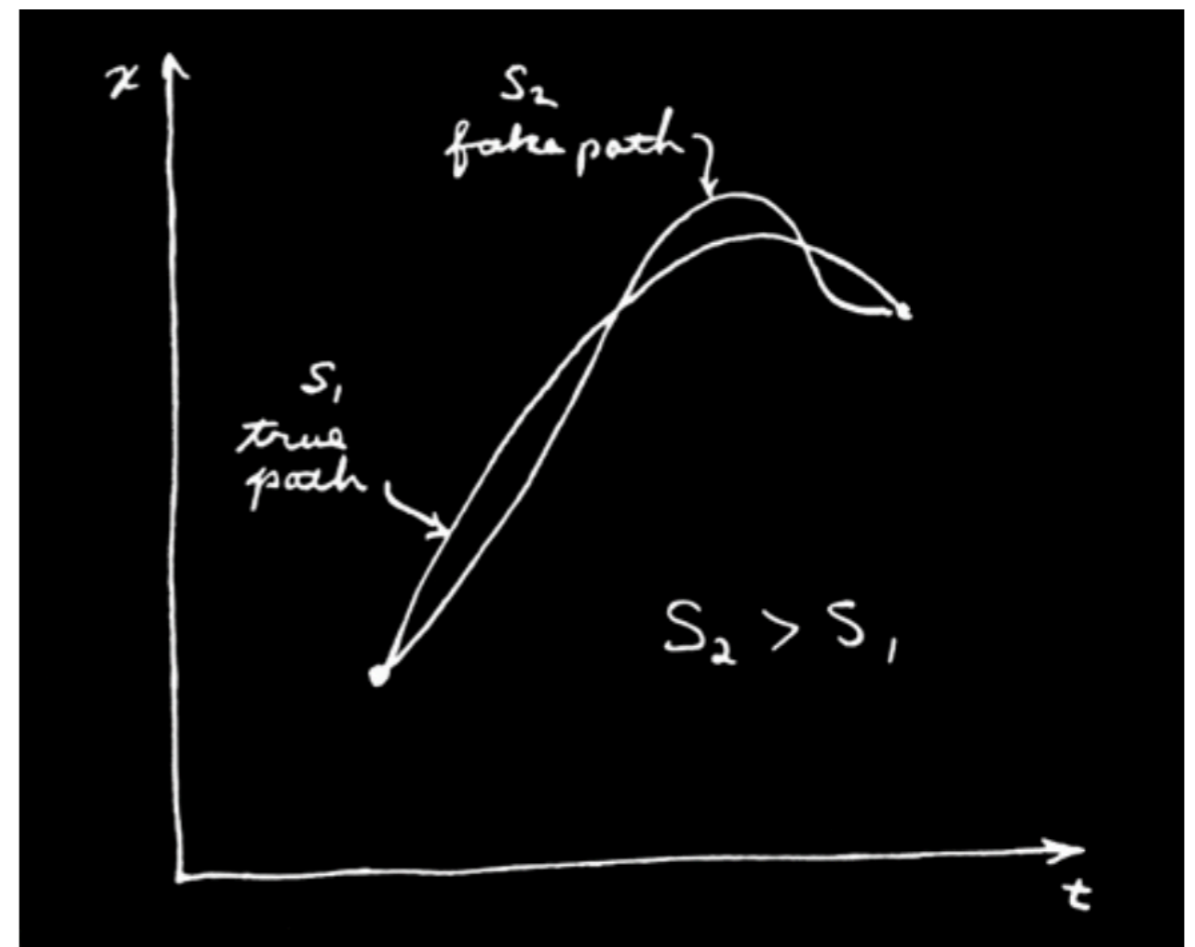


Fig. 19-7.

That's a possible way. But we can do it better than that. When we have a quantity which has a minimum —for instance, in an ordinary function like the temperature— one of the properties of the minimum is that if we go away from the minimum in the *first* order, the deviation of the function from its minimum value is only *second* order. At any place else on the curve, if we move a small distance the value of the function changes also in the first order. But at a minimum, a tiny motion away makes, in the first approximation, no difference (Fig. 19–8).

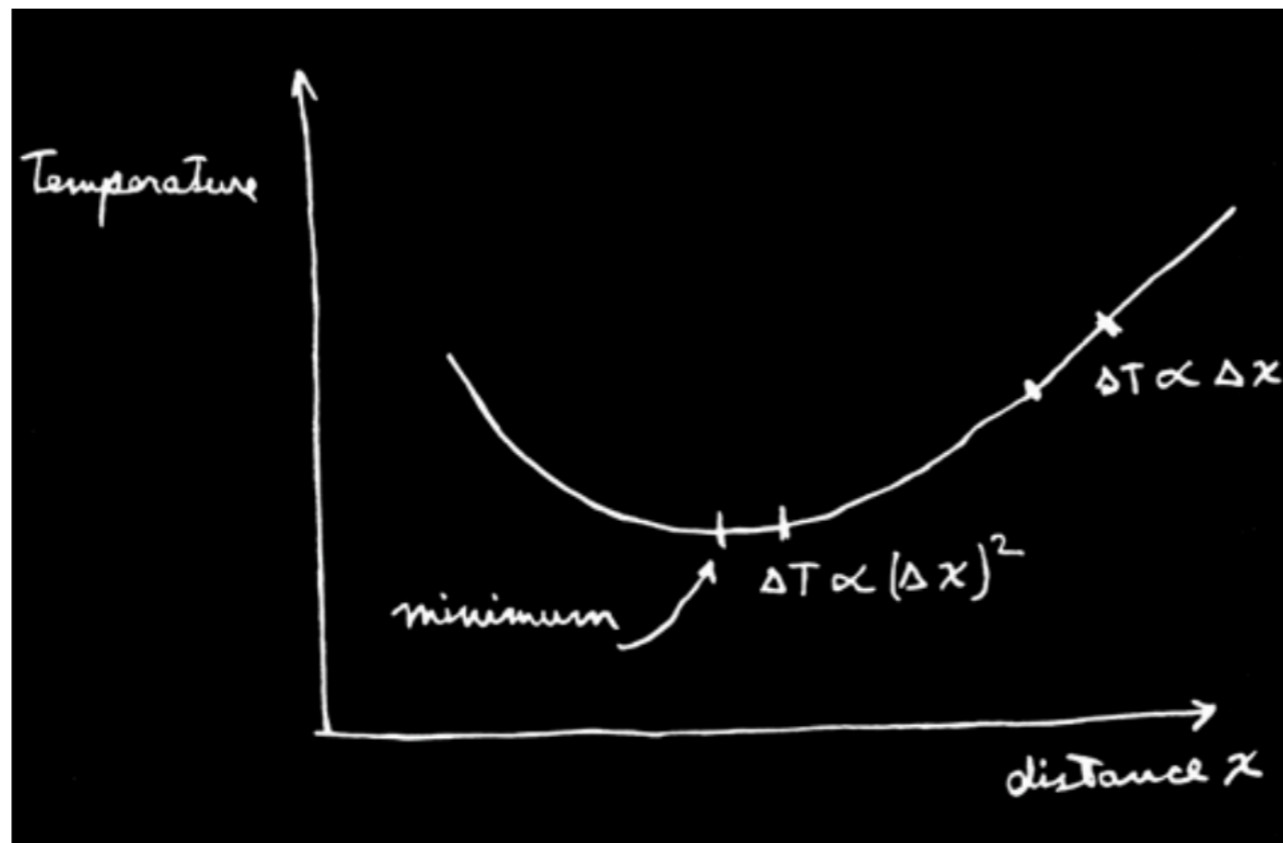


Fig. 19–8.

That is what we are going to use to calculate the true path. If we have the true path, a curve which differs only a little bit from it will, in the first approximation, make no difference in the action. Any difference will be in the second approximation, if we really have a minimum.

That is easy to prove. If there is a change in the first order when I deviate the curve a certain way, there is a change in the action that is *proportional* to the deviation. The change presumably makes the action greater; otherwise we haven't got a minimum. But then if the change is *proportional* to the deviation, reversing the sign of the deviation will make the action less.

We would get the action to increase one way and to decrease the other way. The only way that it could really be a minimum is that in the *first* approximation it doesn't make any change, that the changes are proportional to the square of the deviations from the true path.

So we work it this way: We call $\underline{x(t)}$ (with an underline) the true path—the one we are trying to find. We take some trial path $x(t)$ that differs from the true path by a small amount which we will call $\eta(t)$ (eta of t ; Fig. 19-9).

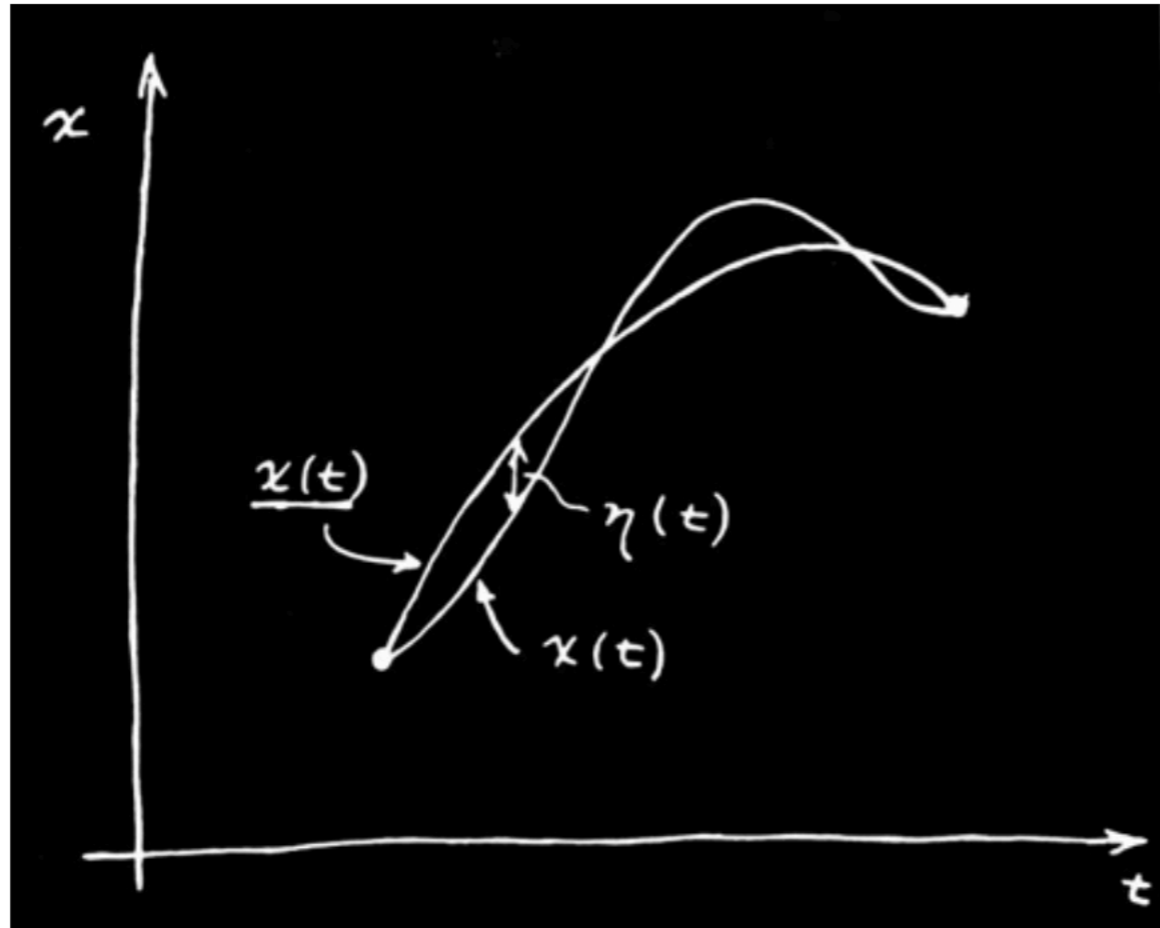


Fig. 19-9.

Now the idea is that if we calculate the action S for the path $x(t)$, then the difference between that S and the action that we calculated for the path $\underline{x(t)}$ —to simplify the writing we can call it \underline{S} —the difference of \underline{S} and S must be zero in the first-order approximation of small η . It can differ in the second order, but in the first order the difference must be zero.

And that must be true for any η at all. Well, not quite. The method doesn't mean anything unless you consider paths which all begin and end at the same two points—each path begins at a certain point at t_1 and ends at a certain other point at t_2 , and those points and times are kept fixed. So the deviations in our η have to be zero at each end, $\eta(t_1) = 0$ and $\eta(t_2) = 0$. With that condition, we have specified our mathematical problem.

If you didn't know any calculus, you might do the same kind of thing to find the minimum of an ordinary function $f(x)$. You could discuss what happens if you take $f(x)$ and add a small amount h to x and argue that the correction to $f(x)$ in the first order in h must be zero at the minimum. You would substitute $x + h$ for x and expand out to the first order in h ... just as we are going to do with η .

The idea is then that we substitute $x(t) = \underline{x}(t) + \eta(t)$ in the formula for the action:

$$S = \int \left[\frac{m}{2} \left(\frac{dx}{dt} \right)^2 - V(x) \right] dt,$$

where I call the potential energy $V(x)$. The derivative dx/dt is, of course, the derivative of $\underline{x}(t)$ plus the derivative of $\eta(t)$, so for the action I get this expression:

$$S = \int_{t_1}^{t_2} \left[\frac{m}{2} \left(\frac{d\underline{x}}{dt} + \frac{d\eta}{dt} \right)^2 - V(\underline{x} + \eta) \right] dt.$$

Now I must write this out in more detail. For the squared term I get

$$\left(\frac{d\underline{x}}{dt} \right)^2 + 2 \frac{d\underline{x}}{dt} \frac{d\eta}{dt} + \left(\frac{d\eta}{dt} \right)^2.$$

But wait. I'm not worrying about higher than the first order, so I will take all the terms which involve η^2 and higher powers and put them in a little box called 'second and higher order.' From this term I get only second order, but there will be more from something else. So the kinetic energy part is

$$\frac{m}{2} \left(\frac{d\underline{x}}{dt} \right)^2 + m \frac{d\underline{x}}{dt} \frac{d\eta}{dt} + (\text{second and higher order}).$$

Now we need the potential V at $\underline{x} + \eta$. I consider η small, so I can write $V(x)$ as a Taylor series. It is approximately $V(\underline{x})$; in the next approximation (from the ordinary nature of derivatives) the correction is η times the rate of change of V with respect to \underline{x} , and so on:

$$V(\underline{x} + \eta) = V(\underline{x}) + \eta V'(\underline{x}) + \frac{\eta^2}{2} V''(\underline{x}) + \dots$$

I have written V' for the derivative of V with respect to x in order to save writing. The term in η^2 and the ones beyond fall into the 'second and higher order' category and we don't have to worry about them. Putting it all together,

$$S = \int_{t_1}^{t_2} \left[\frac{m}{2} \left(\frac{d\underline{x}}{dt} \right)^2 - V(\underline{x}) + m \frac{d\underline{x}}{dt} \frac{d\eta}{dt} - \eta V'(\underline{x}) + (\text{second and higher order}) \right] dt.$$

Now if we look carefully at the thing, we see that the first two terms which I have arranged here correspond to the action \underline{S} that I would have calculated with the true path \underline{x} . The thing I want to concentrate on is the change in S — the difference between the S and the \underline{S} that we would get for the right path. This difference we will write as δS , called the variation in S .

Leaving out the ‘second and higher order’ terms, I have for δS

$$\delta S = \int_{t_1}^{t_2} \left[m \frac{d\underline{x}}{dt} \frac{d\eta}{dt} - \eta V'(\underline{x}) \right] dt.$$

Now the problem is this: Here is a certain integral. I don’t know what the \underline{x} is yet, but I do know that *no matter what η is*, this integral must be zero. Well, you think, the only way that that can happen is that what multiplies η must be zero. But what about the first term with $d\eta/dt$? Well, after all, if η can be anything at all, its derivative is anything also, so you conclude that the coefficient of $d\eta/dt$ must also be zero. That isn’t quite right. It isn’t quite right because there is a connection between η and its derivative; they are not absolutely independent, because $\eta(t)$ must be zero at both t_1 and t_2 .

The method of solving all problems in the calculus of variations always uses the same general principle. You make the shift in the thing you want to vary (as we did by adding η); you look at the first-order terms; then you always arrange things in such a form that you get an integral of the form ‘some kind of stuff times the shift (η),’ but with no other derivatives (no $d\eta/dt$). It must be rearranged so it is always ‘something’ times η . You will see the great value of that in a minute. (There are formulas that tell you how to do this in some cases without actually calculating, but they are not general enough to be worth bothering about; the best way is to calculate it out this way.)

How can I rearrange the term in $d\eta/dt$ to make it have an η ? I can do that by integrating by parts. It turns out that the whole trick of the calculus of variations consists of writing down the variation of S and then integrating by parts so that the derivatives of η disappear. It is always the same in every problem in which derivatives appear.

You remember the general principle for integrating by parts. If you have any function f times $d\eta/dt$ integrated with respect to t , you write down the derivative of ηf :

$$\frac{d}{dt}(\eta f) = \eta \frac{df}{dt} + f \frac{d\eta}{dt}.$$

The integral you want is over the last term, so

$$\int f \frac{d\eta}{dt} dt = \eta f - \int \eta \frac{df}{dt} dt.$$

In our formula for δS , the function f is m times $d\underline{x}/dt$; therefore, I have the following formula for δS .

$$\begin{aligned} \delta S = m \frac{d\underline{x}}{dt} \eta(t) \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt} \left(m \frac{d\underline{x}}{dt} \right) \eta(t) dt \\ - \int_{t_1}^{t_2} V'(\underline{x}) \eta(t) dt. \end{aligned}$$

The first term must be evaluated at the two limits t_1 and t_2 . Then I must have the integral from the rest of the integration by parts. The last term is brought down without change.

Now comes something which always happens—the integrated part disappears. (In fact, if the integrated part does not disappear, you restate the principle, adding conditions to make sure it does!) We have already said that η must be zero at both ends of the path, because the principle is that the action is a minimum provided that the varied curve begins and ends at the chosen points.

The condition is that $\eta(t_1) = 0$, and $\eta(t_2) = 0$. So the integrated term is zero. We collect the other terms together and obtain this:

$$\delta S = \int_{t_1}^{t_2} \left[-m \frac{d^2 \underline{x}}{dt^2} - V'(\underline{x}) \right] \eta(t) dt.$$

The variation in S is now the way we wanted it—there is the stuff in brackets, say F , all multiplied by and integrated from t_1 to t_2 .

We have that an integral of something or other times $\eta(t)$ is always zero:

$$\int F(t) \eta(t) dt = 0.$$

I have some function of t ; I multiply it by $\eta(t)$; and I integrate it from one end to the other. And no matter what the η is, I get zero. That means that the function $F(t)$ is zero. That's obvious, but anyway I'll show you one kind of proof.

Suppose that for $\eta(t)$ I took something which was zero for all t except right near one particular value. It stays zero until it gets to this t , then it blips up for a moment and blips right back down (Fig. 19–10).

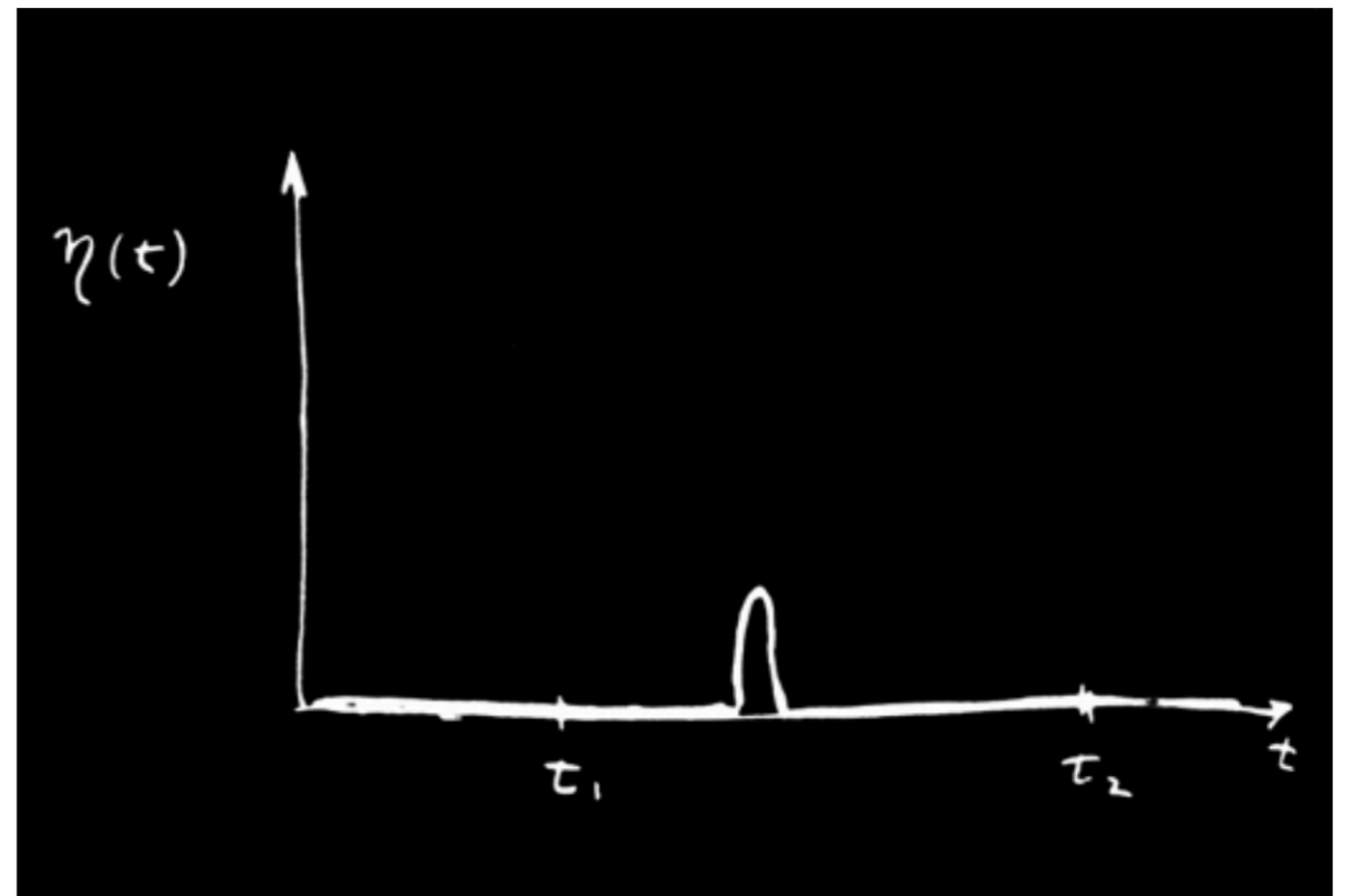


Fig. 19–10.

When we do the integral of this η times any function F , the only place that you get anything other than zero was where $\eta(t)$ was blipping, and then you get the value of F at that place times the integral over the blip. The integral over the blip alone isn't zero, but when multiplied by F it has to be; so the function F has to be zero where the blip was. But the blip was anywhere I wanted to put it, so F must be zero everywhere.

We see that if our integral is zero for any η , then the coefficient of η must be zero. The action integral will be a minimum for the path that satisfies this complicated differential equation:

$$\left[-m \frac{d^2 \underline{x}}{dt^2} - V'(\underline{x}) \right] = 0.$$

It's not really so complicated; you have seen it before. It is just $F = ma$. The first term is the mass times acceleration, and the second is the derivative of the potential energy, which is the force.

So, for a conservative system at least, we have demonstrated that the principle of least action gives the right answer; it says that the path that has the minimum action is the one satisfying Newton's law.

One remark: I did not prove it was a *minimum*—maybe it's a maximum. In fact, it doesn't really have to be a minimum. It is quite analogous to what we found for the 'principle of least time' which we discussed in optics. There also, we said at first it was 'least' time. It turned out, however, that there were situations in which it wasn't the *least* time. The fundamental principle was that for any *first-order variation* away from the optical path, the change in time was zero; it is the same story. What we really mean by 'least' is that the first-order change in the value of S , when you change the path, is zero. It is not necessarily a 'minimum.'

Next, I remark on some generalizations. In the first place, the thing can be done in three dimensions. Instead of just x , I would have x , y and z as functions of t ; the action is more complicated. For three-dimensional motion, you have to use the complete kinetic energy— $(m/2)$ times the whole velocity squared. That is

$$\text{KE} = \frac{m}{2} \left[\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 + \left(\frac{dz}{dt} \right)^2 \right].$$

Also, the potential energy is a function of x , y and z . And what about the path? The path is some general curve in space, which is not so easily drawn, but the idea is the same. And what about the η ? Well, can have three components. You could shift the paths in x , or in y , or in z —or you could shift in all three directions simultaneously. So η would be a vector. This doesn't really complicate things too much, though. Since only the first-order variation has to be zero, we can do the calculation by three successive shifts. We can shift η only in the x -direction and say that coefficient must be zero. We get one equation. Then we shift it in the y -direction and get another. And in the z -direction and get another. Or, of course, in any order that you want. Anyway, you get three equations. And, of course, Newton's law is really three equations in the three dimensions—one for each component. I think that you can practically see that it is bound to work, but we will leave you to show for yourself that it will work for three dimensions. Incidentally, you could use any coordinate system you want, polar or otherwise, and get Newton's laws appropriate to that system right off by seeing what happens if you have the shift η in radius, or in angle, etc.

Similarly, the method can be generalized to any number of particles. If you have, say, two particles with a force between them, so that there is a mutual potential energy, then you just add the kinetic energy of both particles and take the potential energy of the mutual interaction. And what do you vary? You vary the paths of *both* particles. Then, for two particles moving in three dimensions, there are six equations.

You can vary the position of particle 1 in the x -direction, in the y -direction, and in the z -direction, and similarly for particle 2; so there are six equations. And that's as it should be. There are the three equations that determine the acceleration of particle 1 in terms of the force on it and three for the acceleration of particle 2, from the force on it. You follow the same game through, and you get Newton's law in three dimensions for any number of particles.

I have been saying that we get Newton's law. That is not quite true, because Newton's law includes nonconservative forces like friction. Newton said that is equal to any F . But the principle of least action only works for *conservative* systems—where all forces can be gotten from a potential function. You know, however, that on a microscopic level—on the deepest level of physics—there are no nonconservative forces. Nonconservative forces, like friction, appear only because we neglect microscopic complications—there are just too many particles to analyze. But the *fundamental* laws *can* be put in the form of a principle of least action.

Let me generalize still further. Suppose we ask what happens if the particle moves relativistically. We did not get the right relativistic equation of motion; $F = ma$ is only right nonrelativistically. The question is: Is there a corresponding principle of least action for the relativistic case? There is. The formula in the case of relativity is the following:

$$S = -m_0 c^2 \int_{t_1}^{t_2} \sqrt{1 - v^2/c^2} dt - q \int_{t_1}^{t_2} [\phi(x, y, z, t) - \mathbf{v} \cdot \mathbf{A}(x, y, z, t)] dt.$$

The first part of the action integral is the rest mass times times the integral of a function of velocity, $\sqrt{1 - v^2/c^2}$

Then instead of just the potential energy, we have an integral over the scalar potential ϕ and over \mathbf{v} times the vector potential \mathbf{A} . Of course, we are then including only electromagnetic forces. All electric and magnetic fields are given in terms of ϕ and \mathbf{A} . This action function gives the complete theory of relativistic motion of a single particle in an electromagnetic field.

Of course, wherever I have written \mathbf{v} , you understand that before you try to figure anything out, you must substitute dx/dt for v_x and so on for the other components. Also, you put the point along the path at time t , $x(t)$, $y(t)$, $z(t)$ where I wrote simply x , y , z . Properly, it is only after you have made those replacements for the \mathbf{v} 's that you have the formula for the action for a relativistic particle. I will leave to the more ingenious of you the problem to demonstrate that this action formula does, in fact, give the correct equations of motion for relativity. May I suggest you do it first without the \mathbf{A} , that is, for no magnetic field? Then you should get the components of the equation of motion, $d\mathbf{p}/dt = -q\nabla\phi$, where, you remember, $\mathbf{p} = m_0\mathbf{v}/\sqrt{1 - v^2/c^2}$

It is much more difficult to include also the case with a vector potential. The variations get much more complicated. But in the end, the force term does come out equal to $q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$, as it should. But I will leave that for you to play with.

I would like to emphasize that in the general case, for instance in the relativistic formula, the action integrand no longer has the form of the kinetic energy minus the potential energy. That's only true in the nonrelativistic approximation. For example, the term $m_0c^2\sqrt{1 - v^2/c^2}$ is not what we have called the kinetic energy. The question of what the action should be for any particular case must be determined by some kind of trial and error. It is just the same problem as determining what are the laws of motion in the first place. You just have to fiddle around with the equations that you know and see if you can get them into the form of the principle of least action.

One other point on terminology. The function that is integrated over time to get the action S is called the *Lagrangian*, \mathcal{L} , which is a function only of the velocities and positions of particles. So the principle of least action is also written

$$S = \int_{t_1}^{t_2} \mathcal{L}(x_i, v_i) dt,$$

where by x_i and v_i are meant all the components of the positions and velocities. So if you hear someone talking about the ‘Lagrangian,’ you know they are talking about the function that is used to find S . For relativistic motion in an electromagnetic field

$$\mathcal{L} = -m_0c^2 \sqrt{1 - v^2/c^2} - q(\phi - \mathbf{v} \cdot \mathbf{A}).$$

Also, I should say that S is not really called the ‘action’ by the most precise and pedantic people. It is called ‘Hamilton’s first principal function.’ Now I hate to give a lecture on ‘the-principle-of-least-Hamilton’s-first-principal-function.’ So I call it ‘the action.’ Also, more and more people are calling it the action. You see, historically something else which is not quite as useful was called the action, but I think it’s more sensible to change to a newer definition. So now you too will call the new function the action, and pretty soon everybody will call it by that simple name.

Now I want to say some things on this subject which are similar to the discussions I gave about the principle of least time. There is quite a difference in the characteristic of a law which says a certain integral from one place to another is a minimum—which tells something about the whole path—and of a law which says that as you go along, there is a force that makes it accelerate. The second way tells how you inch your way along the path, and the other is a grand statement about the whole path. In the case of light, we talked about the connection of these two.

Now, I would like to explain why it is true that there are differential laws when there is a least action principle of this kind. The reason is the following: Consider the actual path in space and time. As before, let's take only one dimension, so we can plot the graph of x as a function of t . Along the true path, S is a minimum. Let's suppose that we have the true path and that it goes through some point a in space and time, and also through another nearby point b (Fig. 19-11).

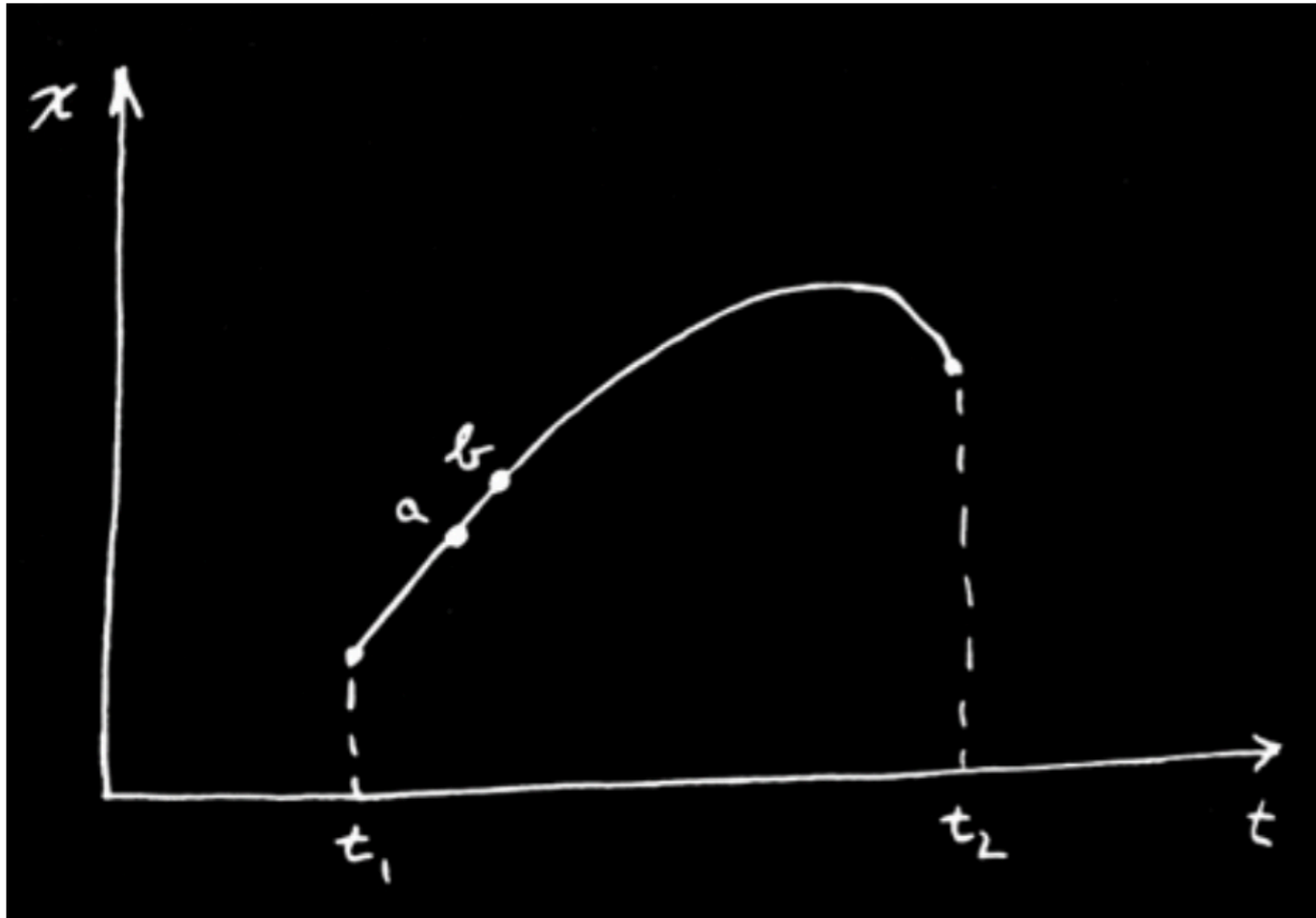


Fig. 19-11.

Now if the entire integral from t_1 to t_2 is a minimum, it is also necessary that the integral along the little section from a to b is also a minimum. It can't be that the part from a to b is a little bit more. Otherwise you could just fiddle with just that piece of the path and make the whole integral a little lower. So every subsection of the path must also be a minimum. And this is true no matter how short the subsection. Therefore, the principle that the whole path gives a minimum can be stated also by saying that an infinitesimal section of path also has a curve such that it has a minimum action.

Now if we take a short enough section of path—between two points a and b and very close together—how the potential varies from one place to another far away is not the important thing, because you are staying almost in the same place over the whole little piece of the path. The only thing that you have to discuss is the first-order change in the potential. The answer can only depend on the derivative of the potential and not on the potential everywhere.

So the statement about the gross property of the whole path becomes a statement of what happens for a short section of the path— a differential statement. And this differential statement only involves the derivatives of the potential, that is, the force at a point. That's the qualitative explanation of the relation between the gross law and the differential law.

In the case of light we also discussed the question: How does the particle find the right path? From the differential point of view, it is easy to understand. Every moment it gets an acceleration and knows only what to do at that instant. But all your instincts on cause and effect go haywire when you say that the particle decides to take the path that is going to give the minimum action. Does it 'smell' the neighboring paths to find out whether or not they have more action? In the case of light, when we put blocks in the way so that the photons could not test all the paths, we found that they couldn't figure out which way to go, and we had the phenomenon of diffraction.

Is the same thing true in mechanics? Is it true that the particle doesn't just 'take the right path' but that it looks at all the other possible trajectories? And if by having things in the way, we don't let it look, that we will get an analog of diffraction? The miracle of it all is, of course, that it does just that. That's what the laws of quantum mechanics say. So our principle of least action is incompletely stated. It isn't that a particle takes the path of least action but that it smells all the paths in the neighborhood and chooses the one that has the least action by a method analogous to the one by which light chose the shortest time. You remember that the way light chose the shortest time was this: If it went on a path that took a different amount of time, it would arrive at a different phase. And the total amplitude at some point is the sum of contributions of amplitude for all the different ways the light can arrive. All the paths that give wildly different phases don't add up to anything. But if you can find a whole sequence of paths which have phases almost all the same, then the little contributions will add up and you get a reasonable total amplitude to arrive. The important path becomes the one for which there are many nearby paths which give the same phase.

It is just exactly the same thing for quantum mechanics. The complete quantum mechanics (for the nonrelativistic case and neglecting electron spin) works as follows: The probability that a particle starting at point 1 at the time t_1 will arrive at point 2 at the time t_2 is the square of a probability amplitude. The total amplitude can be written as the sum of the amplitudes for each possible path—for each way of arrival. For every $x(t)$ that we could have—for every possible imaginary trajectory—we have to calculate an amplitude. Then we add them all together. What do we take for the amplitude for each path? Our action integral tells us what the amplitude for a single path ought to be. The amplitude is proportional to some constant times $e^{iS/\hbar}$, where S is the action for that path. That is, if we represent the phase of the amplitude by a complex number, the phase angle is S/\hbar . The action S/\hbar has dimensions of energy times time, and Planck's constant \hbar has the same dimensions. It is the constant that determines when quantum mechanics is important.

Here is how it works: Suppose that for all paths, S is very large compared to \hbar . One path contributes a certain amplitude. For a nearby path, the phase is quite different, because with an enormous S even a small change in means a completely different phase—because \hbar is so tiny. So nearby paths will normally cancel their effects out in taking the sum—except for one region, and that is when a path and a nearby path all give the same phase in the first approximation (more precisely, the same action within \hbar). Only those paths will be the important ones. So in the limiting case in which Planck's constant \hbar goes to zero, the correct quantum-mechanical laws can be summarized by simply saying: 'Forget about all these probability amplitudes. The particle does go on a special path, namely, that one for which S does not vary in the first approximation.' That's the relation between the principle of least action and quantum mechanics. The fact that quantum mechanics can be formulated in this way was discovered in 1942 by a student of that same teacher, Bader, I spoke of at the beginning of this lecture.

Quantum mechanics was originally formulated by giving a differential equation for the amplitude (Schrödinger) and also by some other matrix mathematics (Heisenberg).]

Now I want to talk about other minimum principles in physics. There are many very interesting ones. I will not try to list them all now but will only describe one more. Later on, when we come to a physical phenomenon which has a nice minimum principle, I will tell about it then. I want now to show that we can describe electrostatics, not by giving a differential equation for the field, but by saying that a certain integral is a maximum or a minimum. First, let's take the case where the charge density is known everywhere, and the problem is to find the potential ϕ everywhere in space. You know that the answer should be

$$\nabla^2 \phi = -\rho/\epsilon_0.$$

But another way of stating the same thing is this: Calculate the integral U^* , where

$$U^* = \frac{\epsilon_0}{2} \int (\nabla \phi)^2 dV - \int \rho \phi dV,$$

which is a volume integral to be taken over all space. This thing is a minimum for the correct potential distribution $\phi(x, y, z)$.

We can show that the two statements about electrostatics are equivalent. Let's suppose that we pick any function ϕ . We want to show that when we take for ϕ the correct potential $\underline{\phi}$, plus a small deviation f , then in the first order, the change in U^* is zero. So we write

$$\phi = \underline{\phi} + f.$$

The $\underline{\phi}$ is what we are looking for, but we are making a variation of it to find what it has to be so that the variation of U^* is zero to first order. For the first part of U^* , we need

$$(\nabla\phi)^2 = (\nabla\underline{\phi})^2 + 2\nabla\underline{\phi} \cdot \nabla f + (\nabla f)^2.$$

The only first-order term that will vary is

$$2\nabla\underline{\phi} \cdot \nabla f.$$

In the second term of the quantity U^* , the integrand is

$$\rho\phi = \rho\underline{\phi} + \rho f,$$

whose variable part is ρf . So, keeping only the variable parts, we need the integral

$$\Delta U^* = \int (\epsilon_0 \nabla\underline{\phi} \cdot \nabla f - \rho f) dV.$$

Now, following the old general rule, we have to get the darn thing all clear of derivatives of f . Let's look at what the derivatives are. The dot product is

$$\frac{\partial\underline{\phi}}{\partial x} \frac{\partial f}{\partial x} + \frac{\partial\underline{\phi}}{\partial y} \frac{\partial f}{\partial y} + \frac{\partial\underline{\phi}}{\partial z} \frac{\partial f}{\partial z},$$

which we have to integrate with respect to x , to y , and to z . Now here is the trick: to get rid of $\partial f/\partial x$ we integrate by parts with respect to x . That will carry the derivative over onto the $\underline{\phi}$. It's the same general idea we used to get rid of derivatives with respect to t . We use the equality

$$\int \frac{\partial \phi}{\partial x} \frac{\partial f}{\partial x} dx = f \frac{\partial \phi}{\partial x} - \int f \frac{\partial^2 \phi}{\partial x^2} dx.$$

The integrated term is zero, since we have to make f zero at infinity. (That corresponds to making η zero at t_1 and t_2 . So our principle should be more accurately stated: U^* is less for the true ϕ than for any other $\phi(x, y, z)$ having the same values at infinity.) Then we do the same thing for y and z . So our integral ΔU^* is

$$\Delta U^* = \int (-\epsilon_0 \nabla^2 \underline{\phi} - \rho) f dV.$$

In order for this variation to be zero for any f , no matter what the coefficient of f must be zero and, therefore,

$$\nabla^2 \underline{\phi} = -\rho/\epsilon_0.$$

We get back our old equation. So our ‘minimum’ proposition is correct.

We can generalize our proposition if we do our algebra in a little different way. Let’s go back and do our integration by parts without taking components. We start by looking at the following equality:

$$\nabla \cdot (f \nabla \underline{\phi}) = \nabla f \cdot \nabla \underline{\phi} + f \nabla^2 \underline{\phi}.$$

If I differentiate out the left-hand side, I can show that it is just equal to the right-hand side. Now we can use this equation to integrate by parts. In our integral ΔU^* , we replace $\nabla \underline{\phi} \cdot \nabla f$ by $\nabla \cdot (f \nabla \underline{\phi}) - f \nabla^2 \underline{\phi}$, which gets integrated over volume.

The divergence term integrated over volume can be replaced by a surface integral:

$$\int \nabla \cdot (f \nabla \underline{\phi}) dV = \int f \nabla \underline{\phi} \cdot \underline{n} da.$$

Since we are integrating over all space, the surface over which we are integrating is at infinity. There, f is zero and we get the same answer as before.

Only now we see how to solve a problem when we *don't* know where all the charges are. Suppose that we have conductors with charges spread out on them in some way. We can still use our minimum principle if the potentials of all the conductors are fixed. We carry out the integral for U^* only in the space outside of all conductors. Then, since we can't vary $\underline{\phi}$ on the conductor, f is zero on all those surfaces, and the surface integral

$$\int f \nabla \underline{\phi} \cdot \underline{n} da$$

is still zero. The remaining volume integral

$$\Delta U^* = \int (-\epsilon_0 \nabla^2 \underline{\phi} - \rho) f dV$$

is only to be carried out in the spaces between conductors. Of course, we get Poisson's equation again,

$$\nabla^2 \underline{\phi} = -\rho/\epsilon_0.$$

So we have shown that our original integral U^* is also a minimum if we evaluate it over the space outside of conductors all at fixed potentials (that is, such that any trial $\phi(x, y, z)$ must equal the given potential of the conductors when (x, y, z) is a point on the surface of a conductor).

There is an interesting case when the only charges are on conductors. Then

$$U^* = \frac{\epsilon_0}{2} \int (\nabla \phi)^2 dV.$$

Our minimum principle says that in the case where there are conductors set at certain given potentials, the potential between them adjusts itself so that integral U^* is least. What is this integral? The term $\nabla \phi$ is the electric field, so the integral is the electrostatic energy. The true field is the one, of all those coming from the gradient of a potential, with the minimum total energy.

I would like to use this result to calculate something particular to show you that these things are really quite practical. Suppose I take two conductors in the form of a cylindrical condenser (Fig. 19-12).

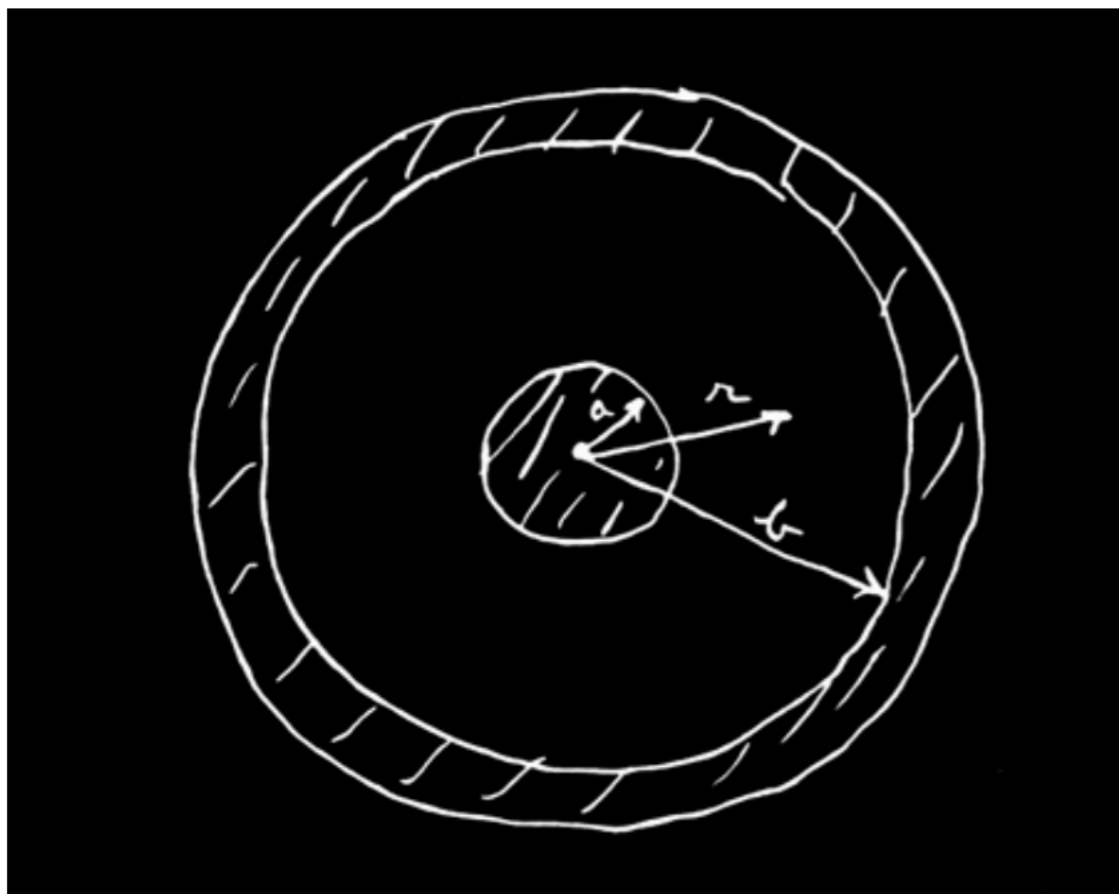


Fig. 19-12.

The inside conductor has the potential V , and the outside is at the potential zero. Let the radius of the inside conductor be a and that of the outside b . Now we can suppose *any* distribution of potential between the two. If we use the *correct* ϕ , and calculate $\epsilon_0/2 \int (\nabla \phi)^2 dV$, it should be the energy of the system, $\frac{1}{2}CV^2$. So we can also calculate C by our principle. But if we use a wrong distribution of potential and try to calculate the capacity C by this method, we will get a capacity that is too big, since V is specified. Any assumed potential ϕ that is not the exactly correct one will give a fake C that is larger than the correct value. But if my false ϕ is any rough approximation, the C will be a good approximation, because the error in U^* is second order in the error in ϕ .

Suppose I don't know the capacity of a cylindrical condenser. I can use this principle to find it. I just guess at the potential function ϕ until I get the lowest C . Suppose, for instance, I pick a potential that corresponds to a constant field. (You know, of course, that the field isn't really constant here; it varies as $1/r$.) A field which is constant means a potential which goes linearly with distance. To fit the conditions at the two conductors, it must be

$$\phi = V \left(1 - \frac{r - a}{b - a} \right).$$

This function is V at $r = a$, zero at $r = b$, and in between has a constant slope equal to $-V/(b - a)$. So what one does to find the integral U^* is multiply the square of this gradient by $\epsilon_0/2$ and integrate over all volume. Let's do this calculation for a cylinder of unit length. A volume element at the radius r is $2\pi r dr$. Doing the integral, I find that my first try at the capacity gives

$$\frac{1}{2} CV^2 (\text{first try}) = \frac{\epsilon_0}{2} \int_a^b \frac{V^2}{(b - a)^2} 2\pi r dr.$$

The integral is easy; it is just

$$\pi V^2 \left(\frac{b + a}{b - a} \right).$$

So I have a formula for the capacity which is not the true one but is an approximate job:

$$\frac{C}{2\pi\epsilon_0} = \frac{b + a}{2(b - a)}.$$

It is, naturally, different from the correct answer $C = 2\pi\epsilon_0/\ln(b/a)$, but it's not too bad

Let's compare it with the right answer for several values of b/a . I have computed out the answers in Table 19–1.

$\frac{b}{a}$	$\frac{C_{\text{true}}}{2\pi\epsilon_0}$	$\frac{C(\text{first approx.})}{2\pi\epsilon_0}$
2	1.4427	1.500
4	0.721	0.833
10	0.434	0.611
100	0.217	0.51
1.5	2.4663	2.50
1.1	10.492059	10.500000

Even when b/a is as big as 2 — which gives a pretty big variation in the field compared with a linearly varying field — I get a pretty fair approximation. The answer is, of course, a little too high, as expected. The thing gets much worse if you have a tiny wire inside a big cylinder. Then the field has enormous variations and if you represent it by a constant, you're not doing very well. With $b/a = 100$, we're off by nearly a factor of two. Things are much better for small b/a . To take the opposite extreme, when the conductors are not very far apart—say $b/a = 1.1$ — then the constant field is a pretty good approximation, and we get the correct value for C to within a tenth of a percent.

Now I would like to tell you how to improve such a calculation. (Of course, you know the right answer for the cylinder, but the method is the same for some other odd shapes, where you may not know the right answer.) The next step is to try a better approximation to the unknown true ϕ . For example, we might try a constant plus an exponential ϕ , etc. But how do you know when you have a better approximation unless you know the true ϕ ? Answer: You calculate C ; the lowest C is the value nearest the truth.

Let us try this idea out. Suppose that the potential is not linear but say quadratic in r — that the electric field is not constant but linear. The most general quadratic form that fits at and $\phi = 0$ at $r = b$ and $\phi = V$ at $r = a$ is

$$\phi = V \left[1 + \alpha \left(\frac{r - a}{b - a} \right) - (1 + \alpha) \left(\frac{r - a}{b - a} \right)^2 \right],$$

where α is any constant number. This formula is a little more complicated. It involves a quadratic term in the potential as well as a linear term. It is very easy to get the field out of it. The field is just

$$E = -\frac{d\phi}{dr} = -\frac{\alpha V}{b - a} + 2(1 + \alpha) \frac{(r - a)V}{(b - a)^2}.$$

Now we have to square this and integrate over volume. But wait a moment. What should I take for α ? I can take a parabola for the ϕ ; but what parabola? Here's what I do: Calculate the capacity with an *arbitrary* α . What I get is

$$\frac{C}{2\pi\epsilon_0} = \frac{a}{b - a} \left[\frac{b}{a} \left(\frac{\alpha^2}{6} + \frac{2\alpha}{3} + 1 \right) + \frac{1}{6} \alpha^2 + \frac{1}{3} \right].$$

It looks a little complicated, but it comes out of integrating the square of the field. Now I can pick my α . I know that the truth lies lower than anything that I am going to calculate, so whatever I put in for α is going to give me an answer too big. But if I keep playing with α and get the lowest possible value I can, that lowest value is nearer to the truth than any other value. So what I do next is to pick the α that gives the minimum value for C .

Working it out by ordinary calculus, I get that the minimum C occurs for $\alpha = -2b/(b + a)$. Substituting that value into the formula, I obtain for the minimum capacity

$$\frac{C}{2\pi\epsilon_0} = \frac{b^2 + 4ab + a^2}{3(b^2 - a^2)}.$$

I've worked out what this formula gives for C various values of b/a . I call these numbers $C(\text{quadratic})$. Table 19–2 compares $C(\text{quadratic})$ with the true C .

$\frac{b}{a}$	$\frac{C_{\text{true}}}{2\pi\epsilon_0}$	$\frac{C(\text{quadratic})}{2\pi\epsilon_0}$
2	1.4427	1.444
4	0.721	0.733
10	0.434	0.475
100	0.217	0.347
1.5	2.4663	2.4667
1.1	10.492059	10.492063

For example, when the ratio of the radii is 2 to 1, I have 1.444, which is a very good approximation to the true answer, 1.4427. Even for larger b/a , it stays pretty good—it is much, much better than the first approximation. It is even fairly good—only off by 10 percent—when b/a is 10 to 1. But when it gets to be 100 to 1—well, things begin to go wild. I get that C is 0.347 instead of 0.217. On the other hand, for a ratio of radii of 1.5, the answer is excellent; and for b/a of 1.1, the answer comes out 10.492063 instead of 10.492059. Where the answer should be good, it is very, very good.

I have given these examples, first, to show the theoretical value of the principles of minimum action and minimum principles in general and, second, to show their practical utility—not just to calculate a capacity when we already know the answer. For any other shape, you can guess an approximate field with some unknown parameters like α and adjust them to get a minimum. You will get excellent numerical results for otherwise intractable problems.