The Usefulness of Lagrange Multipliers in Everyday Life

Abstract

Lagrange multipliers are a very useful technique in multivariable calculus, but all too often they are poorly taught and poorly understood. With luck, this overview will help to make the concept and its applications a bit clearer.One of the most common problems in calculus is that of finding maxima or minima (in general, "extrema") of function, but it is often difficult to find a closed form for the function being extremized. Such difficulties often arise when one wishes to maximize or minimize a function subject to fixed outside conditions or constraints. The method of Lagrange multipliers is a powerful tool for solving this class of problems without the need to explicitly solve the conditions and use them to eliminate extra variables.

Introduction

Put more simply, its usually not enough to ask, "How do I minimize the aluminium needed to make this can?" (The answer to that is clearly "Make a really, really small can!") You need to ask, "How do I minimize the aluminium *whilemaking sure the can will hold 10 liters of milk*?" or similarly, "How do I maximize my factory's profit given that I only have $15,000 to invest?" Or to take a more

sophisticated example, "How quickly will the roller coaster reach the ground assuming it stays on the track?" In general, Lagrange multipliers are useful when some of the variables in the simplest description of a problem are made redundant by the constraints.

A Classic Example: the "milkmaid problem"
To give a specific, intuitive illustration of this kind of problem, we will consider a classic example which I believe is known as the "Milkmaid Problem". It can be phrased as follows

M $g(x,y) = 0$

C
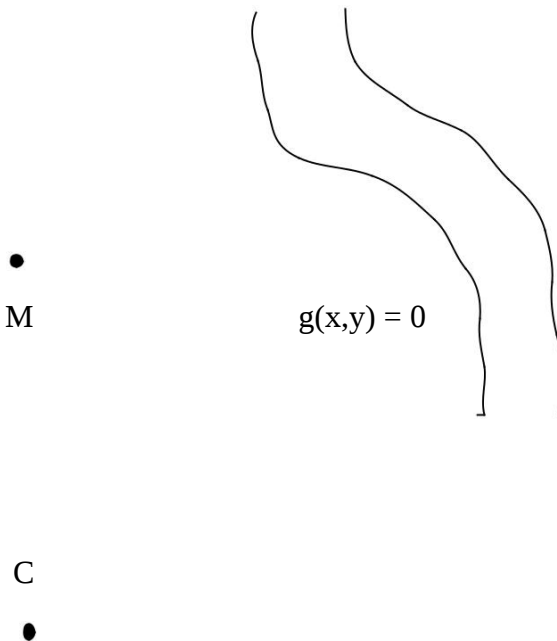
Its milking time at the farm and the milk maid has been sent to the field to get the day's milk. She's in a hurry to get back for a date with a handsome young goatherd, so

she wants to finish her job as quickly as possible. However, before she can gather the milk, she has to rinse out her bucket in the nearby river.

Just when she reaches point **M,** our heroine spots the cow, way down at point **C.** Because she is in a hurry, she wants to take the shortest possible path from where she is to the river and then to the cow. If the near bank of the river is a curve satisfying the function(x,y)=0, what is the shortest path for the milkmaid to take? (To keep things simple, we assume that the field is flat and uniform and that all points on the river bank are equally good.)

To put this into more mathematical terms, the milkmaid wants to find the point **P** for which the distance d (**M,P**) from **M** to **P** plus the distance d(**P,C**) from **P** to **C** is a minimum (we assume that the field is flat, so a straight line is the shortest distance between two points). It's not quite for that matter **P** anywhere on the line between **M** and **C**) we have to impose the constraint that **P** is a point on the riverbank. Formally, we must minimize the function f(**P**)=d(**M,P**) + d(**P,C**), subject to the constraint that g(**P**) = 0

Graphical Inspiration for the Method
Our first way thinking about this problem can be obtained directly from the picture itself. We'll use an obscure fact from geometry; for every point **P** on a given ellipse, the total distance from one focus of the ellipse to **P** and then to the other focus is exactly the same. (You don't need to know where this fact comes from to understand the example! But you can see it work for yourself by drawing a near-perfect ellipse with the help of two nails, a pencil and a loop of string.)

In our problem, that means that the milkmaid could get to the cow by way of any point on a giveellipse in the same amount of time: the ellipse are curves of constant f(**P**).Therefore, to find the desired point **P** on the river bank, we must simply find the smallest ellipse n that intersects the curve of the river. Just to be clear, only the constant "the f(**P**)" property is really important the fact that these curves are eclipses is just lucky convenience (ellipses are easy to

draw). The same idea will work no matter what space the curves happen to be.
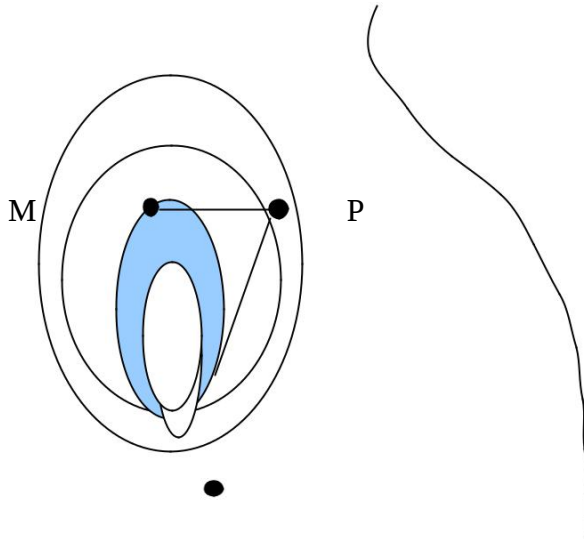
The image at right shows a sequence of ellipses of larger and larger size whose forces are **M** and **C,** ending with the one that is just *tangent* to the riverbank. This is a very significant word! It is obvious from the picture that the "perfect" ellipse and the river are truly tangential to each other at the ideal point **P.** More mathematically, this means that the normal vector to the ellipse is in the same direction as the normal vector to the riverbank. A few minutes' thought about pictures like this will convince you that this fact is not specific to this problem; it is a general property whenever you have constraints. And that is the insight that leads us to the method of Lagrange multipliers.

The Mathematics of Lagrange Multipliers

In multivariable calculus, the gradient of a function h (written

h) is a normal vector to a curve (in two dimensions) or a surface (in higher dimensions) on which h is constant: n=

h (**P**). The length of the normal vector doesn't matter: any constant multiple of

hP is also a normal vector. In our case, we have two functions, whose vectors are parallel,

$$\Delta f(\mathbf{P})=)= f(\mathbf{P}) - \ g (P)$$

The unknown constant multiplier is necessary because the magnitudes of the two gradients may **P** be different. (Remrmber, all we know is that their directions are the same.)

In D dimensions, we now have D+1 equations in D+1 unknowns. D of the unknowns are the coordinates of P (e.g x,y, and z for D=3), and the other is the new unknown constant .. The equation for the gradients derived above is a vector equation, so it proviydes D equations of constraint. I once got stuck on an exam at this point, dont let this happen to you! The original constraint equation g(**P**)=0 is the final equation of the system. Thus, in general, a unique solution exists.

As in many maximum/minimum problems, cases do exist with mulitple solutions. There can even be an infinite number of solutions if the constraints are particularly degenerate; imagine if the milkmaid and the cow were both already standing right at the bank of a straight river, for example. In many cases, the actual value of the

Lagrange multiplier isn't intresting, but there are some situations in which it can give useful information (as discussed below).

That's it; thats all there is to Lagrange multipliers. Just set the gradient of the function you want to extremize equal to the gradient of the constraint function. You'll get a vector's worth of (algebraic)

equations, and together with the original constraint equation they determine the solution.

A formal mathematical inspiration

There is another way to think of Lagrange multipliers that may be more helpful in some situations and that can provide a better way to remember the details of the technique (particularly with multiple constraints as described below). Once again, we start with a function f(**P**) that we wish to extremize a function in multivariable calculus is to set f(**P**)=0. How can we put this condition together with the constraint that we have?

One answer is to add a new variable to the problem, and to define a new function to extremize:

F(**P**,   f(**P**)- g(**P**)

Some references call this f "THE Lagrangian function" I am not familiar with that usage, although it must be related to the somewhat similar "Lagrangian" used in advanced physics.)

We next set

F (**P**,) =0, but keep in mind that the gradient is now D + 1 dimensional: one of its components is a partial derivative with respect to . If you set this new component of the gradient equal to zero, you get the constraint equation g(**P**)=0. Meanwhile, the old components of the gradient treat as a constant, so it just pulls through. Thus the other

D equations are precisely the D equations found in the graphical approach above.

As presented here, this is a just a trick to help you reconstruct the equations you need. However, for those who go on to use Lagrange multipliers in the calculus of variations, this is generally the most

useful approach. I suspect that it is in fact very fundamental; my comments about the meaning of the multiplier below are a step toward exploring it in more depth, but I have never spent the time to work out the details.
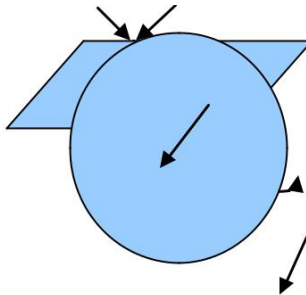

Several Constraints at Once

If you have more than one constraint, all you need to do is to replace the right hand side of the equation with the sum of the gradients of each constraint function, each with its own (different) Lagrange multiplier. This is usually only relevant in at least three dimensions (since two constraints in two dimensions generally intersect at isolated points).

Again, it is easy to understand this graphically. [My thanks to Eric Ojard for suggesting this approach]. Consider the example shown at right: the solution is constrained to lied on the brown plane (as an equation, "g(**P**) =0") and also to lie on the purple ellipsoid ("h(**P**)=0"). For both to be true, the solution must lie on the black ellipse where the two intersect. The important observation is that both normal vectors are perpendicular

to the intersection curve at each point. In fact, any vector perpendicular to it can be written as a linear combination of the two normal vectors. (Assuming the two are linearly independent! If not, the two constraints may already give a specific solution in our example this would happen if the

plane constraint was exactly tangent to the ellipsoid constraint at a single point.)
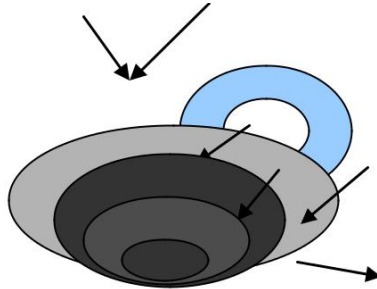
The significance of this becomes clear when we consider a three dimensional analogue of the milk maid problem. The pink ellisoids at right all have the same two foci (which are faintly visible as black dots in the middle), and represent surfaces of constant total distance for travel from one focus to the surface and back to the other. As in two dimensions, the optimal ellipsoid is tangent to the constraint curve and consequently its normal vector is perpendicular to the combined constraint (as shown). Thus the normal vector can be written as a linear combination of the normal vectors of the constraint surfaces. In equations, this statement reads

$$f(\mathbf{P}) = g(\mathbf{P}) + h(\mathbf{P}).$$

just as described above. The generalization to more constraints and higher dimensions is exactly the same.

The meaning of the multiplier

As a final note, I'll say a few words about what the Lagrange multiplier "means" in ways inspired by both

physics and economics. In our mostly geometrical discussion so far, was just an artificial variable that lets us compare the directions of the gradients without worrying about their magnitudes. But in cases where the function

f(**P**) and the constraint g(**P**) have specific meanings, the Lagrange multiplier often has an identifiable significance as well.

One example of this is inspired by the physics of forces and potential energy. In the formal approach based on the combined "Lagrangian function" F(**P**,) described two sections above, the constraint function g(**P**) can be thought of as "competing" with the desired function f(**P**) to "pull" the point **P** to its minimum or maximum. The Lagrange multiplier can be thought of as a measure of how hard g(**P**) has to pull in order to make those "forces" balance out on the constraint surface. (This generalizes naturally to multiple constraints, which typically "pull" in different directions.) And in fact, that word "forces" is very significant: in physics based on Lagrange multipliers in the calculus of variations (as described below) this analogy turns out to be literally true, is the force of constraint.

Application

The Lagrange multiplier has meaning in economics as well. If you're maximizing profit subject to a limited resource, is that resource's marginal value (sometimes called the "shadow price" of the resource). Specifically, the value of the Lagrange Multiplier is the rate at which the optimal value of the function f(**P**) changes if you change the constraint.

To express this mathematically (following the approach of this economics-inspired tutorial by Martin Osborne)

write the constraint in the form "g(**P**)=g (x,y)=c? For some constant c. (This is mathematically equivalent to our usual g(**P**)=0, but allows us to easily describe a whole family of constraints. Also, I am writing this in terms of just two coordinates x and y for clarity, but the generalization to more is straightforward). For any given value of c, we can use Lagrange multipliers to find the optimal value of f(**P**) and the point where it occurs. Call that optimal value fo, occurring at coordinates (x0 , y0 and with Lagrange multiplier . The answers

we get will depend on what value we used for c in the constraint, so we can think of these as functions of c: f0(c), x0 (c) etc.

To find how the optimal value changes when you change the constraint, just take the derivative: df0/dc. Of course, f(**P**) only depends on c because the optimal coordinates (x0 , y0) depend on c: we could write it as f0(x0(c), y0(c). So we have to use the (multi variable) chain rule:

df0/dc=f0 /x0 dx0

In the final step, I've suggestively written this as a dot product between the gradient of f and the derivatives of the coordinate vector. So here's the clever trick: use the lagrange multiplier equation to substitutef= g:

df0/dc=f0 /x0 dx0

But the constraint function is *always* equal to c, so dg/dc=1. Thus, df/dc = . That is, the Lagrange multiplier is the rate of change of the optimal value with respect to changes in the constraint.

This is a powerful result, but be careful when using it! In particular, you have to make sure that your constraint function is written in just the right way. You would get the exact same optimal value whether you wrote "g(x,y)= x + y=0" or "g(x,y)= -2x -2y= 0", but the resulting Lagrange multipliers would be quite different.

Example of Lagrange Multipliier in Action
A box of minimal surface area

What shape should a rectangle box with a specific volume (in three dimensions) be in order to minimize its surface area? (Question like this are very important for business that want to save money on packing materials.) Some people may be able to guess the answer intuitively, but we can prove it using Lagrange multipliers.

Let the lengths of the box's edges be x, y and z. then the constrains of volume is simply g(x, y, z) = xyz-V=0, and the function to minimize is f(x, y, z)=2(xy+xz+yz). The method is straightforward to apply:

2<y+z, x+z, x+y>= f(x,y,z) = g(x,y,z)= <yz, xz, xy>

(The angle bracket notation <a, b, c> is my favorite way to denote a vector). Now just solve those three equational; the solution is x=y=z=4/ . We could eliminate from the problem by using xyz=V, but we don't need to: it is ready clear that the optimal shape is a cube.

The closest approach of a line to a point
This example isn't the perfect illustration of where Lagrange multipliers are useful, since it is fairly easy to solve without them and not all that convenient to solve with them. But it's a very simple idea and because of a dumb mistake on my part it was the first example that I applied the technique to.

When I first took multivariable calculus (and before we learned about Lagrange multipliers) my teacher showed the example of finding the point **P** =<x, y> on a line (y=mx+b) that was closest to a given a point **Q** = <x0 , y0 > The function to minimize is of course

d(**P, Q**) = sqrt[(x-x0)2 =sqrt[(x-x0) + (y-y)]

(Here, "sqrt" means square root, of course; that's hard to draw in plain text).

The teacher went through the problem on the board in the most direct way (I'll explain it later) but it was taking him a while and I was a little bored, so I idly started working the problem myself while he talked. I just leapt in and set $d(x,y) = 0$ so

$$<x-x , y-y >sqrt[(x-x) + y-y = <0,0>$$

and thus x=x and y=y . My mistake here is obvious, so I won't blame you for having a laugh at my expense: I forgot to impose the constraint that <x,y> be on the line! (In my defense, I wasn't really focusing on what I was doing, since I was listening to lecture at the same time)> I felt a little silly, but I didn't think much more about it.

Happily, we learned about Lagrange multipliers the very next week and I immediately saw that my mistake had been a perfect introduction to the technique. We write the equation of the line as g(x,y)= y-mx -b=0, so g(x,y) = <-m,1>. So we just set the to gradients equal (up to the usual factor of),giving

<x-x , y-y >/sqrt[(x-x) + (y-y)]= <-m, 1>

The second component of this equation is just an equation for , so we can substitute that value for into the first component equation. The denominators are the same and cancel, leaving just (x-x)= - m(y-y). Finally, we substitute y=mx+b, giving x-x =-m x-mb + mb
+ my , so we come to the final answer: x=(x + my -mb)/) m + 1). And thus

y=(mx + m y + b)(?(m + 1)

So what did my teacher actually do? He used the equestion of the line to substitute y for x in d(**P, Q**) which left us with an "easy" single variable function to deal with...but a rather complicated on:

d(**p,Q**)=sqrt[(x-x) + (mx + b-Y)]

To solve the problem from the point, you take the derivative and set it equal to zero as usual. It's a bit of pain, since the function is a mess, but the answer is xx=(x+my-mb)/(m + 1.

That's exactly what we got earlier, so both methods seem to work. In

this case. The second method may be a little faster (though I didn't show you all of the work), but in more complicated problems Lagrange multipliers are often much easier than the direct approach.

Lagrange multipliers in the calculus of variations (often in physics). This section will be brief, in part because most readers have probably never heard of the calculus of variations. Many people first see this idea in advanced classes that cover Lagrangian mechanics and that will be the perspective taken here(in particular, I will use names inspired by physics). If you don't already know the basics of this subject (specifically, the Eler-Lagrange equations) you'll probably want to just skip this section.

The calculus of variations is essentially an extension of calculus to the case where the basic variables are not simple numbers x (which can be though of as position) but functions x (which in physics corresponds to a position that changes in time) Rather than seeking the numbers x that extremize a function f(x) we seek the *functions* x(t) that extremize the integral (dt) of a function L[x(t), x(t), t] where x (t) are the time derivatives of x (t). (The reason we have to integrate first is to get an ordinary number out: we know what "minimum" and "minimum" mean for numbers, but there could be any number of definitions of those concepts for functions). In most cases, we integrate between fixed values t and t and we hold the values x(t) and x(t) fixed. In physics that means that the initial and final positions are held

constant, we're intrested finding the"best" path to get between them; L defines what we mean by "best").

The solutions to this problem can be shown to satisfy the Euler-Lagrange equations (I have suppressed the "(t)" in the functions x (t):

(Note that the derivative d/dt is a total derivative, while the derivatives with respect to and are "partials" at least formally.

Imposing constraints on this process is often essential. In physics, it is common for an object to be constrained on some track or surface, or for various coordinates to be related (like position and angle when a wheel rolls without slipping as discussed below). To do this, we follow a simple generalization of the procedure we used in ordinary calculus. First, we write the constraint as a function set equal to zero:g(x, t)=0. (Constraints that necessarily involve the derivatives of x often cannot be solved). And second, we add a term to the function L that is multiplied by a new function (t): L [x,x, t] + (t)g(x,t)

From here, we proceed exactly as you expect: (t) is treated as another coordinate function, just as was treated as an additional coordinate in ordinary calculus. The Euler-Lagrange equations are then written as

This can be generalized to the case of multiple constraints precisely as before, by introducing additional Lagrange multiplier functions like . There are further generalizations possible to cases where the constraints (s) are linear combinations of derivatives of the coordinates (rather than the coordinates themselves) but I won't go into that much detail here.

As mentioned in the calculus section, the meaning of the Lagrange multiplier function in this case is surprisingly well-defined and can be quite useful. It turns out that **Q** = (t)(g/ X) is precisely the force required to impose the constraint g(x ,t) (in th "direction" of x) This is fairly natural: the constraint term (g) added to the Lagrangian plays the same role as a (negative) potential energy -V

constraint, so we can compute the resulting force as(-Vconstr)= g in something reminiscent of the usual way. Thus, for example, Lagrange multipliers can be used to calculate the force you would feel while riding a roller coaster. If you want this information, Lagrange multipliers are one of the best ways to get it.

An example: Rolling without slipping
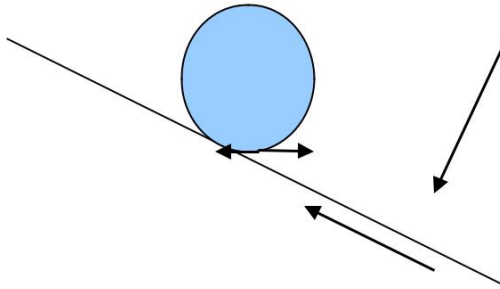
One of the simplest applications of Lagrange multipliers in the calculus of variations is a ball (or other round object) rolling down a slope without slipping in one dimension. (As usual a problem this simple can probably be solved just as easier by other means, but it still illustrates the idea.) Let the ball have mass M, moment of inertia I, and radius R and let the angle of the slope be α. We choose the coordinate x to point up the slope and the coordinate to show rotation in the direction that would naturally go in that same direction, so that the "rolling without slipping" conditions is x=R . Both x and are functions of time, x(t) and (t), but for clarity I will not write that dependence explicitly.

In general, the kinetic energy T of an object undergoing both translational and rotational motion is T=1/2 M(x) +

½ I( ). (As before, the prime(') denotes a time derivative, so this is a function of velocity and angular velocity). Meanwhile, the potential energy V of

the ball can be written as V=Mgh=Mgx sin . Thus, the Lagrangian for the system is

L=T-V=1/2 M(x) + ½ I(θ) – Mgx sin

If we solved the Euler Lagrange equations for this Lagrangian as it stands, we could find that x(t) described the ball sliding down the slope with constant acceleration in the absence of friction while (t) described the ball rotating with constant angular velocity: the rotational motion and translational motion are completely independent (and no torques act on the system). To impose the condition of rolling without slipping, we use a Lagrange multiplier function (t) to force the constraint function G(x, θ,t)=x-R to vanish:

L=1/2 M(x1)2 + ½ I(θ1)2 –Mgx sin + (x-R)

(That is multiplying the constraint function, as usual). We now find the Euler-Lagrange equations for each of the three functions of time: x, θ and . Respectively, the results are:

Mx" + Mg sin α - =0

I θ"+ R + 0, and

X–R θ=0

It is then straight forward to solve for the three "unknowns" in these equations:

X"= -g sin α MR (MR + 1)

θ" = -g sin α MR/ (MR + 1)

=Mg sin α 1/(MR + 1).

The first two equations give the constant acceleration and angular acceleration experienced as the ball rolls down slope. And the

equation for is used in finding the forces that implement the constraint in the problem. Specifically, these are the force and torque due to friction felt by the ball:

$Fx = \partial G/\partial x = Mg \sin \alpha \; I/(MR2 + 1)$

$t = \partial G/\partial \theta = -RM \sin \alpha \; 1/(MR2 + 1)$

Looking at these results, we can see that the force of friction is positive: it points in the +x direction (up the hill) which means that it is slowing down the translational motion. But the torque is negative: It acts in the direction corresponding to rolling down the hill, which means that the speed of rotation increases as the ball rolls down. That is exactly what we would expect!

As a final note, you may be worried that we would get different answers for the forces of constraint if we just normalized the constraint function $G(x, \theta, t)$ differently (for example, if we set $G = 2 x - 2 R\theta$). Happily, that will not end up affecting the final answers at all. A change in normalization for G will lead to a different answer for (e.g. exactly half of what we found above), but the products of and the derivatives of G will remain the same.