Lagrange Multiplier Method for finding Optimums

Typical of the optimum problems you have solved up to now is:

> A rectangular area adjacent to a stone wall is to be enclosed using chain-link fence. Because the wall forms one side of the rectangle, only the three remaining sides need to be fenced. What is the maximum area that can be enclosed using 50 meters of fence?

The way you would set this up is using the two independent variables, L and W for length and width. Then you'd construct the equations:

$$A = LW$$
$$P = 2W + L$$

The first is the area equation and provides us the function for what is to be maximized (i.e., area, A) as a function of length and width. The second is the constraint equation, where $P = 50$ meters. It indicates that the sum of the twice the width with the length must be 50 meters.

To solve this you could solve the constraint equation for L in terms of W and P, then substitute for L into the area equation:

$$L = P - 2W$$
$$A = W(P - 2W) = WP - 2W^2$$

Now take the derivative of A with respect to W, set it to zero, and solve for W:

$$\frac{dA}{dW} = 0 = P - 4W$$
$$W = \frac{P}{4}$$

Finally back-substitute to find L:

$$L = P - 2W = P - 2\frac{P}{4} = \frac{P}{2}$$

Knowing that $P = 50$ meters, you have now established that $L = 25$ meters and $W = 12.5$ meters, resulting in $A = 312.5$ meters^2.

This is quite a serviceable method for solving problems like this, especially when there are only two independent variables (in the case above those are L and W). But if there are any more than two independent variables, it becomes unclear how to apply this method. To solve the more general case where there might be any number of independent variables, the eighteenth century mathematician, Joseph Louis Lagrange, devised a new method using partial derivatives. I will demonstrate it on the problem we just solved.

First we convert the constraint equation to the form, $g(L, W) = 0$. In our particular case, the constraint equation takes the form:

$$2W + L - P = 0$$

Now we combine the function to be optimized with the constraint as follows:

$$f(L, W) = LW + \lambda (2W + L - P)$$

That is we form the function that is the sum of what we want to optimize with λ times the constraint function (note that some textbooks would have you take the difference rather than the sum of LW with $\lambda(2W + L - P)$, but, as you will see later, this doesn't matter). So what is λ? It is an additional independent variable that we have added to ensure that we are able to arrive at a solution (more on that later). Now we take the partial derivatives of f with each of the three independent variables, L and W, as well as the new one we've introduced, λ.

$$\frac{\partial f}{\partial L} = W + \lambda$$
$$\frac{\partial f}{\partial W} = L + 2\lambda$$
$$\frac{\partial f}{\partial \lambda} = 2W + L - P$$

Setting each equation to zero, we solve them simultaneously:

$$0 = W + \lambda$$
$$0 = L + 2\lambda$$
$$0 = 2W + L - P$$

Solving for L using the last equation, we get the same this we did before, $L = P - 2W$. Solving the first equation for λ gives $\lambda = -W$. Substituting both of those results into the middle equation, we get

$$0 = P - 2W - 2W = P - 4W$$

or equivalently $W = \frac{P}{4}$, which is the same result we got using the old method of solution. Indeed the old method might seem easier and certainly more familiar, and for problems involving just two independent variables, there is no reason not to use the old method. But let's try a problem that involves three variables.

> A box is to be constructed out of various materials. The material to be used for the front and back sides costs \$1 per square meter. The material to be used for the left and right sides costs \$2 per square meter. The material to be used for the top and bottom costs \$4 per square meter. What is the maximum volume that can be enclosed for a total material cost of \$192?

If $C = 192$ is cost, then the constraint equation, in the form required for the Lagrange method, is

$$2LH + 4WH + 8LW - C = 0$$

where the three independent variables, L, W, and H, are the length, width, and height of the box respectively. The function to be optimized is volume, V, given by

$$V = LWH$$

So the Lagrange function is

$$f(L, W, H) = LWH + \lambda(2LH + 4WH + 8LW - C)$$

and the partial derivative equations are

$$\frac{\partial f}{\partial L} = 0 = WH + \lambda(2H + 8W)$$

$$\frac{\partial f}{\partial W} = 0 = LH + \lambda(4H + 8L)$$

$$\frac{\partial f}{\partial H} = 0 = LW + \lambda(2L + 4W)$$

$$\frac{\partial f}{\partial \lambda} = 0 = 2LH + 4WH + 8LW - C$$

At first solving this looks daunting – that is until you find the key. By solving the first equation for H and the third for L, you find:

$$H = \frac{-8\lambda W}{W + 2\lambda} \quad \text{and} \quad L = \frac{-4\lambda W}{W + 2\lambda}$$

hence $H = 2L$. Substituting that into the second equation:

$$0 = 2L^2 + \lambda(8L + 8L) = 2L^2 + 16\lambda L$$

so $\lambda = \frac{-L}{8}$. Substituting both those results into the first equation:

$$0 = 2WL - \frac{L}{8}(4L + 8W)$$

yields $W = \frac{L}{2}$. At this point we have the independent variables, H, W, and λ all in terms of L. Substituting H and W into the last equation:

$$0 = 4L^2 + 4L^2 + 4L^2 - C$$

or $L = \sqrt{\frac{C}{12}}$. Back-substituting to find W and H, you find that $W = \sqrt{\frac{C}{48}}$ and $H = \sqrt{\frac{C}{3}}$. Putting in $C = 192$ and taking the product of the three dimensions, we find that $V = 64$ meters^3.

So why does it work and what is λ? Imagine you are hiking along a trail that traverses the side of a valley. The side of the valley is a surface, so it takes two coordinates, x and y (i.e., longitude and latitude), to identify a point on that surface. But park-service rules constrain you to hike only along the trail. Every point in the valley has an altitude, $h(x, y)$. So where along the trail does your altitude reach a local maximum or a local minimum?

At every point in the valley you can assign a direction in the x-y plane along which the valley slopes up at the greatest rate. And if you were to walk in that direction, you can assign a magnitude of the up-slope you would experience. The combination of this direction and magnitude forms a vector that is known as the *gradient* of the altitude function, which is denoted ∇h. If $\mathbf{i}$ is the unit vector in the x direction and $\mathbf{j}$ is the unit vector in the y direction, then

$$\nabla h = \frac{\partial h}{\partial x}\mathbf{i} + \frac{\partial h}{\partial y}\mathbf{j}$$

Back to the hiking-trail analogy: Observe that whenever the trail is at a local maximum or local minimum, *the the trail runs exactly perpendicular to the gradient vector of the altitude.* Think about this carefully, because it is at the heart of the Lagrange multipliers method. Indeed, what the Lagrange method does is to solve for the point (or points) where the trail is perpendicular to the gradient vector.

If the trail is defined by a constraint function, $g(x, y) = 0$, we can also take the gradient vector, ∇g, of that constraint function. Furthermore the direction of ∇g will always be exactly perpendicular to the trail. Hence the Lagrange method solves for the point (or points) along the trail where ∇g is parallel to ∇h. For two vectors to be parallel, they must be scalar multiples of each other. So Lagrange solves for $\nabla h = -\lambda \nabla g$, where λ is some nonzero real number.

The Lagrange method forms the function, $f(x, y) = h(x, y) + \lambda g(x, y)$, where h is the valley's altitude function, g is the trail's constraint function, and $-\lambda$ is the unknown scalar that equates ∇h and ∇g at the solution point(s). That is, at a solution point, (x, y), you will have

$$\frac{\partial h}{\partial x}\mathbf{i} + \frac{\partial h}{\partial y}\mathbf{j} \;=\; -\lambda\left(\frac{\partial g}{\partial x}\mathbf{i} + \frac{\partial g}{\partial y}\mathbf{j}\right)$$

Recall that earlier I mentioned that it doesn't matter whether you take the sum or the difference. Now you can see why. Taking the difference, $f(x, y) = h(x, y) - \lambda g(x, y)$, has no effect other than to flip the sign of λ. Also notice that the constraint, $g(x, y) = 0$, is not at all different from the condition, $\lambda g(x, y) = 0$, for any nonzero λ, positive or negative.

Lagrange solves for

$$0 \;=\; \nabla f \;=\; \nabla h + \lambda \nabla g \;=\; \frac{\partial h}{\partial x}\mathbf{i} + \frac{\partial h}{\partial y}\mathbf{j} + \lambda\frac{\partial g}{\partial x}\mathbf{i} + \lambda\frac{\partial g}{\partial y}\mathbf{j}$$

Separating terms according to whether they are multiplied by **i** or **j** yields the two equations

$$0 \;=\; \frac{\partial h}{\partial x} + \lambda\frac{\partial g}{\partial x} \quad \text{and} \quad 0 \;=\; \frac{\partial h}{\partial y} + \lambda\frac{\partial g}{\partial y}$$

But this is still not enough information to arrive at a solution for unknowns. Why? Because we have only two equations, but three unknowns, x, y, and λ. We need one more equation in order to have a path to a solution. Most textbooks tell you to take the partial derivative, $\frac{\partial f}{\partial \lambda}$, and set that to zero, which is what I did in the example problems. But notice that $\frac{\partial f}{\partial \lambda} = 0$ *always yields the original constraint equation.*

$$\frac{\partial f}{\partial \lambda} \;=\; \frac{\partial\,(h + \lambda g)}{\partial \lambda} = g$$

This is because h, the function to be optimized, is independent of λ, hence its partial with repect to λ is zero. So besides solving for the two gradients being parallel, we also require that the solution meet original the constraint condition, which it had to all along. Going back to our trail analogy, we are simply saying, find the point or points where the gradient of the altitude is parallel to the gradient of the constraint function *and* simultaneously satisfy the constraint equation (that is, it's a point *on* the trail).

When you use Lagrange to optimize a function of three independent variables, as we did with volume in the box problem, the three independent variables define a three dimensional space. In this case, the constraint equation represents a surface in that space. The gradient of the constraint function will be a vector that is perpendicular to that surface. The gradient of the function to be optimized (like volume in the box equation) will also be a vector. Again Lagrange solves for the points where the two gradient vectors are parallel and that simultaneously satisfy the constraint equation. The method works just as well when there are more than three independent variables.