Lagrange multipliers

Normally if we want to maximize or minimize a function of two variables $f(x,y)$, then we set

$$\frac{\partial f}{\partial x} = 0, \qquad \frac{\partial f}{\partial y} = 0, \tag{1}$$

solve the two simultaneous equations we get, and we're done. But suppose we have in addition a constraint that says that x and y can only take certain values. We want to maximize (or minimize) the function subject to that constraint. In other words, out of the set of values x, y that satisfy the constraint, we want to find the ones that give the largest value of $f(x,y)$.

All possible constraints can always be rearranged to read $g(x,y) = 0$. For instance, the constraint $x^2y = 3$ and be written $x^2y - 3 = 0$, and so forth. The equation $g(x,y) = 0$ gives us a line or curve in the xy plane. We want to go along that line and find the point on it with the largest value of $f(x,y)$. When we find that largest value, we know that $f(x,y)$ will be *stationary as we move along the line*. That's what being a maximum means. This in turn means that the gradient ∇f is perpendicular to the line of the constraint at this point, because the gradient is always perpendicular to the contour (the line of constant f).

But we know something else also. The constraint is also constant long the line. Of course it is—it's zero by construction $g(x,y) = 0$, and zero is a constant. So it also has a gradient ∇g that is perpendicular to the line. Thus the two gradients aren't equal, but they do point in the same direction. In other words, they are proportional to each other $\nabla f = \lambda \nabla g$, where λ is some number that we don't at present know the value of. This number λ is called a **Lagrange multiplier**. Rearranging, we see that $\nabla f - \lambda \nabla g = 0$, or

$$\nabla F = 0, \qquad \text{where } F(x,y) = f(x,y) - \lambda g(x,y). \tag{2}$$

But if $\nabla F = 0$, then all partial derivatives of F are zero.

This gives us a method for finding our maximum. We simply construct the function $F(x,y)$ as above, but multiplying g by a constant λ, and then we set all partial derivatives of F to zero and solve the set of simultaneous equations we get. This gives us a solution for the maximum of F.

The only problem is that the solution still contains the constant λ, whose value we don't know. But we can solve that problem easily enough. We still have one more equation—the constraint. If we substitute our solution for x and y back into the constraint, we get another equation that we can solve for λ, and then we're done.

Example: Find the maximum of $1 - x^2 - 2y^2$ subject to the constraint that $x + y = 1$.

Our two basic functions are $f(x,y) = 1 - x^2 - 2y^2$ and $g(x,y) = x + y - 1$. Notice that we have rearranged the constraint so that $g(x,y) = 0$. Then the Lagrange function $F(x,y)$ is

$$F(x,y) = f(x,y) - \lambda g(x,y) = 1 - x^2 - 2y^2 - \lambda(x + y - 1). \tag{3}$$

The our simultaneous equations for the solution are

$$\frac{\partial F}{\partial x} = 0 = -2x - \lambda, \qquad \frac{\partial F}{\partial y} = 0 = -4y - \lambda. \tag{4}$$

Hence, the maximum of $f(x, y)$ is at $x = -\frac{1}{2}\lambda$, $y = -\frac{1}{4}\lambda$.

To complete the solution, we just need to know what the value of λ is. To get this, we substitute our x and y into the constraint, to get

$$x + y - 1 = -\frac{1}{2}\lambda - \frac{1}{4}\lambda - 1 = 0, \tag{5}$$

which gives us $\lambda = -\frac{4}{3}$.

Putting it all together, we then find that

$$x = -\frac{\lambda}{2} = \frac{2}{3}, \qquad y = -\frac{\lambda}{4} = \frac{1}{3}. \tag{6}$$

More than two variables: It easy to show that the method generalizes to more than two variables and more than one constraint. If we have a set $\{x_i\}, i = 1 \ldots n$ of variables and a set $g_j(\{x_i\}), j = 1 \ldots m$ of constraints, then one constructs the function

$$F(\{x_i\}) = f(\{x_i\}) - \sum_{j=1}^{m} \lambda_j g_j(\{x_i\}), \tag{7}$$

where $\{\lambda_j\}$ is a set of m unknown Lagrange multipliers. Then we solve the n simultaneous equations

$$\frac{\partial F}{\partial x_i} = 0, \tag{8}$$

to get x_i in terms of the λ_j, and we substitute the answer back into the m constraints to get m more equations that we solve for λ_j.

This is what you need to do for maximizing the entropy, where the variables x_i are now $p(s)$ for different states, and the constraints are given in the question. Evaluating the Lagrange multipliers in that case gives you complicated expressions in terms of the energy and so forth, but it's OK just to leave the multipliers in there. By convention we always write our expressions in terms of the multipliers themselves (temperature, partition function, etc.) rather than energy and so forth—it's more convenient that way.