Lagrange Multipliers

Let us say we have a function f of n variables, i. e., $f \equiv f(x_1, x_2, ..., x_n)$. By doing a Taylor expansion up to first order, the difference in the value of f between two neighboring points, $(x_1, x_2, ..., x_n)$ and $(x_1 + dx_1, x_2 + dx_2, ..., x_n + dx_n)$ can be written as,

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i. \tag{1}$$

If we want to maximize (or minimize) f then we look for a point, $(x_1, x_2, ..., x_n)$ such that $df = 0$ around it. Since, in this case, there is no conditions (constraints) on x_i's all dx_i (and also all x_i) are linearly independent. Therefore, the condition $df = 0$ *necessarily* implies that,

$$\frac{\partial f}{\partial x_i} = 0 \tag{2}$$

for all i.

However, in some cases we encounter problems of maximization (or minimization) with one or more constraints. For example, let us consider the problem of maximization (minimization) of $f(x_1, x_2, ..., x_n)$ with the constraint that another function $\phi(x_1, x_2, ..., x_n) = 0$. In this case, at the point of maximum (or minimum) we still have $df = 0$, but in this case, we cant say that the condition $df = 0$ implies,

$$\frac{\partial f}{\partial x_i} = 0 \tag{3}$$

for all i. This is just because, the condition $\phi(x_1, x_2, ..., x_n) = 0$ renders one of the x_i's (and hence corresponding dx_i) **dependent** on the rest $n - 1$ variables which can be treated as independent [1]. Since not all variables now are independent our arguments that led to Eqn 2 do not hold in this case.

Can we find out a *necessary condition* someway similar to Eqn. 2 for the case of constrained maximization (minimization)? Below, we shall derive such a condition.

As in the preceding case, we consider the problem of maximizing $f(x_1, x_2, ..., x_n)$ with the constraint that $\phi(x_1, x_2, ..., x_n) = 0$.

Since, $\phi = 0$ we have,

$$d\phi = \sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i} dx_i = 0. \tag{4}$$

As discussed earlier, since the condition $df = 0$ holds even for the constrained problem, using above equation we can write, $df + \lambda d\phi = 0$ where λ can take arbitrary value. Now, using Eqn. 1 we obtain,

$$\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} + \lambda \frac{\partial \phi}{\partial x_i} \right) dx_i = 0. \tag{5}$$

We know that not all dx_i's in above equation is independent. In particular, one of them can be treated to be dependent on the rest $(n - 1)$ independent variables. Let we treat, x_n to be dependent on the first

[1] Since $\phi(x_1, x_2, ..., x_n) = 0$ we can write $x_n = h(x_1, x_2, ..., x_{n-2}, x_{n-1})$, where h is a function of the first $n - 1$ variables.

$(n-1)$ linearly independent variables. Let us separate the n^{th} term from the first $(n-1)$ terms in the above equation as follows,

$$\sum_{i=1}^{n-1}\left(\frac{\partial f}{\partial x_i}+\lambda\frac{\partial \phi}{\partial x_i}\right)dx_i+\left(\frac{\partial f}{\partial x_n}+\lambda\frac{\partial \phi}{\partial x_n}\right)dx_n=0. \tag{6}$$

But here λ is arbitrary. So we can choose λ such that the coefficient of dx_n becomes zero. This is actually the most important step. Since the coefficient of dx_n is 0 it does not matter if we treat dx_n in above equation as linearly independent or dependent on the rest of the $n-1$ linearly independent variables – i.e., in effect, dx_n has lost is presence from the above equation. Therefore, by enforcing, $\frac{\partial f}{\partial x_n}+\lambda\frac{\partial \phi}{\partial x_n}=0$ we obtain,

$$\sum_{i=1}^{n-1}\left(\frac{\partial f}{\partial x_i}+\lambda\frac{\partial \phi}{\partial x_i}\right)dx_i=0\,. \tag{7}$$

where all of the dx_i's are linearly independent. Hence we must have, $\left(\frac{\partial f}{\partial x_i}+\lambda\frac{\partial \phi}{\partial x_i}\right)$ for $i=1,2,3,...,n-1$. Combining with the case for $i=n$, we obtain,

$$\left(\frac{\partial f}{\partial x_i}+\lambda\frac{\partial \phi}{\partial x_i}\right)=0\,, \tag{8}$$

for all i. Using vector notation we can rewrite above equation as $\nabla_n f=-\lambda\nabla_n\phi$, where ∇_n denotes the n dimensional gradient operator. λ is called Lagrange's multiplier. This equation gives *necessary condition* for maximum of f subject to the constraint, $\phi=0$.

If we have more than one constraint, namely, $\phi_j(x_1,x_2,..,x_n)=0$, where $j=1,2,...,m$ with $m<n$, then one gets $n-m$ linearly independent dx_i and m dependent dx_i. By introducing m Lagrange multipliers we can set the coefficients of m dependent dx_i to zero. The rest $(n-m)$ coefficients then automatically become zero from the condition of linear independence. The condition for maximum of f is then modified as $\nabla_n f=-\lambda_1\nabla_n\phi_1-\lambda_2\nabla_n\phi_2-......-\lambda_m\nabla_n\phi_m$