N. LAGRANGE MULTIPLIERS METHOD

Imagine a Cartesian coordinate system of $n+m$ dimensions with the axes labelled $x_1, x_2, \ldots, x_{n+m}$ and a function[1] $E(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n+m})$. Suppose that we are interested in finding the lowest value of E, but only among such $\boldsymbol{x}$ that satisfy m conditions (*conditional extremum*): conditional extremum

$$W_i(\boldsymbol{x}) = 0 \tag{N.1}$$

for $i = 1, 2, \ldots, m$. The constraints cause the number of *independent* variables to be n.

If we calculated the differential $\mathrm{d}E$ at point $\boldsymbol{x}_0$, which corresponds to an extremum of E, then we obtain 0:

$$0 = \sum_{j=1}^{n+m} \left(\frac{\partial E}{\partial x_j}\right)_0 \mathrm{d}x_j, \tag{N.2}$$

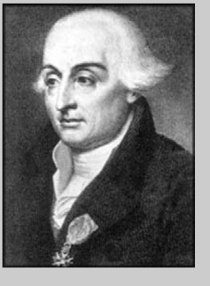
where the derivatives are calculated at the point of the extremum. The quantities $\mathrm{d}x_j$ stand for infinitesimally small increments. From (N.2) we cannot draw the conclusion that the $(\frac{\partial E}{\partial x_j})_0$ are equal to 0. This would be true if the increments $\mathrm{d}x_j$ were independent, but they are not. Indeed, we find the relations between them by making differentials of conditions W_i:

$$\sum_{j=1}^{n+m} \left(\frac{\partial W_i}{\partial x_j}\right)_0 \mathrm{d}x_j = 0 \tag{N.3}$$

for $i = 1, 2, \ldots, m$ (the derivatives are calculated for the extremum).

This means that the number of truly independent increments is only n. Let us try to exploit this. To this end let us multiply each equation (N.3) by a number ϵ_i (*Lagrange multiplier*), which will be fixed in a moment. Then, let us add together all the conditions (N.3), and subtract the result from eq. (N.2). We get

Joseph Louis de Lagrange (1736–1813), French mathematician of Italian origin, self-taught; professor at the Artillery School in Turin, then at the École Normale Supérieure in Paris. His main achievements are in variational calculus, mechanics, and also in number theory, algebra and mathematical analysis.



[1] Symbol E is chosen to suggest that, in our applications, the quantity will have the meaning of energy.

$$\sum_{j=1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

where the summation extends over $n+m$ terms. The summation may be carried out in two steps. First, let us sum up the first n terms, and afterwards sum the other terms

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j+\sum_{j=n+1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0.$$

The multipliers ϵ_i have so far been treated as "undetermined". Well, we may force them to make each of the terms in the second summation equal zero[2]

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0,\quad \text{for } j=n+1,\ldots,n+m. \tag{N.4}$$

Hence, the first summation alone is 0

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

which means that now we have only n increments $\mathrm{d}x_j$, and therefore they are *independent*. Since for any (small) $\mathrm{d}x_j$, the sum is always 0, the only reason for this could be that each parenthesis [ ] *individually equals zero*

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0\quad \text{for } j=1,\ldots,n.$$

Euler equation

This set of n equations (called the *Euler equations*) together with the m conditions (N.1) and m equations (N.4), gives a set of $n+2m$ equations with $n+2m$ unknowns (m epsilons and $n+m$ components x_i of the vector $\boldsymbol{x}_0$).

> For a conditional extremum, the constraint $W_i(\boldsymbol{x})=0$ has to be satisfied for $i=1,2,\ldots,m$ and
>
> $$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0\quad \text{for } j=1,\ldots,n+m.$$
>
> The x_i found from these equations determine the position $\boldsymbol{x}_0$ of the conditional extremum E.

Whether it is a minimum, a maximum or a saddle point, is decisive for the analysis of the matrix of the second derivative (Hessian). If its eigenvalues calculated at $\boldsymbol{x}_0$ are all positive (negative), it is a minimum[3] (maximum), in other cases it is a saddle point.

[2]This is possible if the determinant which is built of coefficients $(\frac{\partial W_i}{\partial x_j})_0$ is non-zero (this is what we have to assume). For example, if several conditions were identical, the determinant would be zero.

[3]In this way we find a minimum; no information is available as to whether is it global or local.

Example 1. ***Minimizing a paraboloid going along a straight line off centre.*** Let us take a paraboloid

$$E(x, y) = x^2 + y^2.$$

This function has, of course, a minimum at $(0, 0)$, but the minimum is of no interest to us. What we want to find is a minimum of E, but only *when* x and y satisfy some conditions. In our case there will only be one:

$$W = \frac{1}{2}x - \frac{3}{2} - y = 0. \tag{N.5}$$

This means that we are interested in a minimum of E *along the straight line* $y = \frac{1}{2}x - \frac{3}{2}$.

The Lagrange multipliers method works as follows:

- We differentiate W and multiply by an unknown (Lagrange) multiplier ϵ thus getting: $\epsilon(\frac{1}{2}\,\mathrm{d}x - \mathrm{d}y) = 0$.
- This result (i.e. 0) is subtracted[4] from $\mathrm{d}E = 2x\,\mathrm{d}x + 2y\,\mathrm{d}y = 0$ and we obtain $\mathrm{d}E = 2x\,\mathrm{d}x + 2y\,\mathrm{d}y - \frac{1}{2}\varepsilon\,\mathrm{d}x + \varepsilon\,\mathrm{d}y = 0$.
- In the last expression, the coefficients at $\mathrm{d}x$ and $\mathrm{d}y$ have to be equal to zero.[5] In this way we obtain two equations: $2x - \frac{1}{2}\epsilon = 0$ and $2y + \epsilon = 0$.
- The third equation needed is the constraint $y = \frac{1}{2}x - \frac{3}{2}$.
- The solution to these three equations gives a set of x, y, ϵ which corresponds to an extremum. We obtain: $x = \frac{3}{5}$, $y = -\frac{6}{5}$, $\epsilon = \frac{12}{5}$. Thus we have obtained, not only the position of the minimum ($x = \frac{3}{5}$, $y = -\frac{6}{5}$), but also the Lagrange multiplier ϵ. The minimum value of E, which has been encountered along the straight line $y = \frac{1}{2}x - \frac{3}{2}$ is equal to $E(\frac{3}{5}, -\frac{6}{5}) = (\frac{3}{5})^2 + (-\frac{6}{5})^2 = \frac{9+36}{25} = \frac{9}{5}$.

Example 2. ***Minimizing a paraboloid moving along a circle (off centre).*** Let us take the same paraboloid (N.5), but put another constraint

$$W = (x - 1)^2 + y^2 - 1 = 0. \tag{N.6}$$

This condition means that we want to go around a circle of radius 1, centred at $(1, 0)$, and see at which point (x, y) we have the lowest value[6] of E. The example is chosen in such a way as to answer the question first *without any calculations*. Indeed, the circle goes through $(0, 0)$, therefore, this point has to be found as the minimum. Beside this, we should find a maximum at $(2, 0)$, because this is the point on the circle which is most distant from $(0, 0)$.

Well, let us see whether the Lagrange multipliers method will give the same result.

After differentiation of W, multiplying it by the multiplier ϵ, subtracting the result from dE and rearranging the terms, we obtain

[4] Or added – no matter (in that case we get a different value of ϵ).

[5] This is only possible now.

[6] Or, in other words, we intersect the paraboloid with the cylindrical surface of radius 1 and the cylinder axis (parallel to the axis of symmetry of the paraboloid) is shifted to $(1, 0)$.

$$\mathrm{d}E = \left[2x - \epsilon(2x-2)\right]\mathrm{d}x + 2y(1-\epsilon)\,\mathrm{d}y = 0,$$

which (after forcing the coefficients at dx and dy to be zero) gives a set of three equations

$$\begin{aligned} 2x - \epsilon(2x-2) &= 0, \\ 2y(1-\epsilon) &= 0, \\ (x-1)^2 + y^2 &= 1. \end{aligned}$$

Please check that this set has the following solutions: $(x, y, \epsilon) = (0, 0, 0)$ and $(x, y, \epsilon) = (2, 0, 2)$. The solution $(x, y) = (0, 0)$ corresponds to the minimum, while the solution $(x, y) = (2, 0)$ corresponds to the maximum.[7] This is what we expected to get.

Example 3. ***Minimizing the mean value of the harmonic oscillator Hamiltonian.*** This example is different: it will pertain to the extremum of a *functional*.[8] *We are often going to encounter this in the methods of quantum chemistry.* Let us take the energy functional

$$E[\phi] = \int_{-\infty}^{\infty} \mathrm{d}x\, \phi^* \hat{H} \phi \equiv \langle \phi | \hat{H} \phi \rangle,$$

where $\hat{H}$ stands for the harmonic oscillator Hamiltonian:

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}kx^2.$$

If we were asked what function ϕ ensures the minimum value of $E[\phi]$, such a function could be found right away, it is $\phi = 0$. Yes, indeed, because the kinetic energy integral and the potential energy integral are positive numbers, except in the situation when $\phi = 0$, then the result is zero. Wait a minute! This is not what we thought of. We want ϕ to have a probabilistic interpretation, like any wave function, and therefore $\langle\phi|\phi\rangle = 1$, and not zero. Well, in such a case we want to minimize $E[\phi]$, *but forcing the normalization condition is satisfied all the time.* Therefore we search for the extremum of $E[\phi]$ *with condition* $W = \langle\phi|\phi\rangle - 1 = 0$. It is easy to foresee that what the method has to produce (if it is to be of any value) is the normalized ground-state wave function for the harmonic oscillator. How will the Lagrange multipliers method get such result?

The answer is on p. 198.

[7] The method does not give us information about the kind of extremum found.

[8] The argument of a functional is a function that produces the value of the functional (a number).