Lagrange Multipliers

1 Introduction

To find maxima and minima of a function of one or more variables it is generally sufficient to equate its first derivatives to zero. Quite often, though, such maxima and minima have to be searched under constraints which arise in relation to the given problem. Consider, for instance, a simple two dimensional gaussian-like surface:

$$z = f(x, y) = \frac{1}{12\pi} \exp\left[-\frac{1}{2}\left(\frac{x^2}{9} + \frac{y^2}{4}\right)\right]$$

To find its maxima and minima we simply have to solve the following system:

$$\begin{cases} \frac{\partial f}{\partial x} = 0 \\ \frac{\partial f}{\partial y} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} -\frac{x}{108\pi} = 0 \\ -\frac{y}{48\pi} = 0 \end{cases}$$

which yields the obvious solution $(x, y) = (0, 0)$. Indeed it is trivial to inspect surface $z = f(x, y)$ and realise that point $(0, 0)$ is a maximum (Figure 1). Things change if we wish to find a maximum for $f(x, y)$, restricted to the surface $y = 1$. It is easy to derive the new position $(0, 1)$ for the maximum in this case. Simply replace $y = 1$ in $z = f(x, y)$, thus obtaining a new function $z = f(x, 1) \equiv g(x)$; the maximum for this function can subsequently found at $x = 0$ and this expression, jointly to the restraint $y = 1$, gives the desired result. Point $(0, 1)$ does not coincide with the *unconstrained maximum* $(0, 0)$ for $f(x, y)$; it is what we call a *constrained maximum* for $f(x, y)$, under the constraint represented by expression $y = 1$.

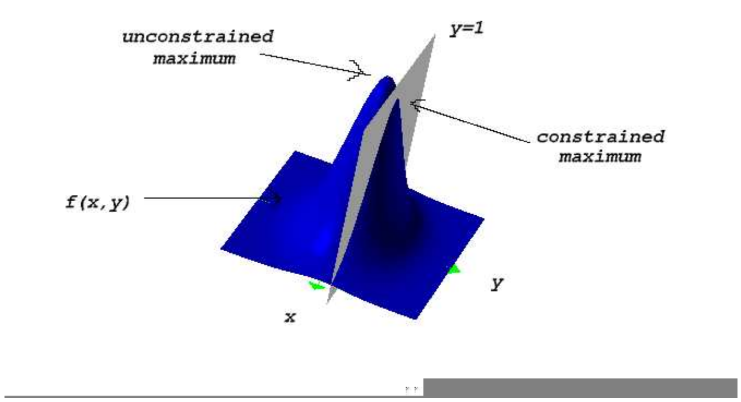

Figure 1: Unconstrained and constrained maximum for function $z = f(x, y) = (1/12\pi)\exp\left[-(1/2)(x^2/9 + y^2/4)\right]$.

James Foadi - Oxford 2011

Lagrange's method of undetermined multipliers is a general technique to deal with the kind of problem just described. Equating first derivatives to zero is no longer sufficient to find maxima and minima under constraints. The constraints introduce relationships among variables which cause some of them to depend on the others.

2 Lagrange's method of undetermined multipliers (theory)

Let us shortly review the conditions according to which a point is an extreme value for a function of one or more variables. We know that maxima and minima are types of extrema; but there are also other types of extrema, like saddle points for instance. In general we can say that, given a function of n variables, $f(x_1, x_2, ..., x_n)$, point $(\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)$ is an *extremum* for f if:

$$\mathrm{d}f \equiv f(\hat{x}_1 + \mathrm{d}x_1, \hat{x}_2 + \mathrm{d}x_2, ..., \hat{x}_n + \mathrm{d}x_n) - f(\hat{x}_1, \hat{x}_2, ..., \hat{x}_n) \approx 0$$

where $(\mathrm{d}x_1, \mathrm{d}x_2, ..., \mathrm{d}x_n)$ is a very small (infinitesimal) arbitrary vector. We can expand $\mathrm{d}f$ using Taylor formula and consider only terms to first degree in $(\mathrm{d}x_1, \mathrm{d}x_2, ..., \mathrm{d}x_n)$. Our condition for an extreme value will then be re-written as:

$$\mathrm{d}f = \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\hat{x}_1} \mathrm{d}x_1 + \left.\frac{\partial f}{\partial x_2}\right|_{x_2=\hat{x}_2} \mathrm{d}x_2 + \cdots + \left.\frac{\partial f}{\partial x_n}\right|_{x_n=\hat{x}_n} \mathrm{d}x_n \approx 0 \tag{1}$$

All $\mathrm{d}x_j$'s are arbitrarily chosen and independent from each other. This means that whatever values we choose for the various $\mathrm{d}x_j$'s, quantity $\mathrm{d}f$ has to equal zero. This can only happen if all individual partial derivatives are equal to zero. We can, thus, state that$(\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)$ is an extremum for f if and only if $\partial f/\partial x_j$ is zero at $x_j = \hat{x}_j$, for all j's:

$$\mathrm{d}f = 0 \quad \Leftrightarrow \quad \left.\frac{\partial f}{\partial x_1}\right|_{x_1=\hat{x}_1} = \left.\frac{\partial f}{\partial x_2}\right|_{x_2=\hat{x}_2} = \cdots = \left.\frac{\partial f}{\partial x_n}\right|_{x_n=\hat{x}_n} = 0 \tag{2}$$

But, what happen if one or more constraints are present? Each constraint can be written as a condition linking the n variables, for instance:

$$\phi(x_1, x_2, ..., x_n) = 0$$

A condition like this is equivalent to making one of the n variables dependent on the other $n-1$ variables. Given that not all variables are independent from each other, condition (2) does not follow any longer from condition (1). How can we behave in such a case? Let us focus on three variables to follow the argument more easily; we will generalize later to n variables.

Consider finding an extremum for $f(x, y, z)$, subject to a constraint represented by the equation $\phi(x, y, z) = 0$. This last equation tells us that one of the variables depends on the other two, e.g. z depends on x and y:

$$\phi(x, y, z) = 0 \quad \Leftrightarrow \quad z = \xi(x, y) \quad \Rightarrow \quad \mathrm{d}z = \frac{\partial \xi}{\partial x}\mathrm{d}x + \frac{\partial \xi}{\partial y}\mathrm{d}y$$

If the extremum is at $(\hat{x}, \hat{y}, \hat{z})$, condition (1) yields, this time:

$$\begin{aligned}
\mathrm{d}f &= \left.\tfrac{\partial f}{\partial x}\right|_{x=\hat{x}} \mathrm{d}x + \left.\tfrac{\partial f}{\partial y}\right|_{y=\hat{y}} \mathrm{d}y + \left.\tfrac{\partial f}{\partial z}\right|_{z=\hat{z}} \mathrm{d}z \\
&= \left.\tfrac{\partial f}{\partial x}\right|_{x=\hat{x}} \mathrm{d}x + \left.\tfrac{\partial f}{\partial y}\right|_{y=\hat{y}} \mathrm{d}y + \left.\tfrac{\partial f}{\partial z}\right|_{z=\hat{z}} \left(\left.\tfrac{\partial \xi}{\partial x}\right|_{x=\hat{x}} \mathrm{d}x + \left.\tfrac{\partial \xi}{\partial y}\right|_{y=\hat{y}} \mathrm{d}y \right) \\
&= \left(\left.\tfrac{\partial f}{\partial x}\right|_{x=\hat{x}} + \left.\tfrac{\partial f}{\partial z}\right|_{z=\hat{z}} \left.\tfrac{\partial \xi}{\partial x}\right|_{x=\hat{x}} \right) \mathrm{d}x + \left(\left.\tfrac{\partial f}{\partial y}\right|_{y=\hat{y}} + \left.\tfrac{\partial f}{\partial z}\right|_{z=\hat{z}} \left.\tfrac{\partial \xi}{\partial y}\right|_{y=\hat{y}} \right) \mathrm{d}y = 0
\end{aligned}$$

James Foadi - Oxford 2011

dx and dy are chosen arbitrarily and independently from each other; consequently the previous condition yields:

$$\left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}} + \left.\frac{\partial f}{\partial z}\right|_{z=\hat{z}} \left.\frac{\partial \xi}{\partial x}\right|_{x=\hat{x}} = 0$$
$$\left.\frac{\partial f}{\partial y}\right|_{y=\hat{y}} + \left.\frac{\partial f}{\partial z}\right|_{z=\hat{z}} \left.\frac{\partial \xi}{\partial y}\right|_{y=\hat{y}} = 0$$

These relations are very different from the simpler ones in (2), and seem to be difficult to solve too. Lagrange had the idea, then, to introduce a further, arbitrary variable λ in the scenario, with the only purpose of making things easier. Consider the following, newly defined, function $F(x, y, z)$, obtained starting from $f(x, y, z)$ and $\phi(x, y, z)$ as:

$$F(x, y, z) \equiv f(x, y, z) + \lambda\phi(x, y, z) \tag{3}$$

(in the definition λ multiplies ϕ, so it is called a *multiplier*). Now, we know that at point $(\hat{x}, \hat{y}, \hat{z})$ both d$f = 0$ and d$\phi = 0$ (this last condition is a consequence of equation $\phi(x, y, z) = 0$). It will be also true, then, that d$F = 0$. From definition (3) we obtain:

$$\mathrm{d}F = \left(\frac{\partial f}{\partial x} + \lambda\frac{\partial \phi}{\partial x}\right)\Big|_{x=\hat{x}}\mathrm{d}x + \left(\frac{\partial f}{\partial y} + \lambda\frac{\partial \phi}{\partial y}\right)\Big|_{y=\hat{y}}\mathrm{d}y + \left(\frac{\partial f}{\partial z} + \lambda\frac{\partial \phi}{\partial z}\right)\Big|_{z=\hat{z}}\mathrm{d}z = 0 \tag{4}$$

dz cannot be chosen independently from dx and dy, but we are free to choose whatever value for the *Lagrange multiplier* λ. We are free, then, to fix it as in the following relation:

$$\left(\frac{\partial f}{\partial z} + \lambda\frac{\partial \phi}{\partial z}\right)\Big|_{z=\hat{z}} = 0 \tag{5}$$

Using this equation in (4), and considering the arbitrariness and independence of dx and dy, we can derive from (4) the two following relations:

$$\left(\frac{\partial f}{\partial x} + \lambda\frac{\partial \phi}{\partial x}\right)\Big|_{x=\hat{x}} = 0 \quad , \quad \left(\frac{\partial f}{\partial y} + \lambda\frac{\partial \phi}{\partial y}\right)\Big|_{y=\hat{y}} = 0 \tag{6}$$

Equations (5) and (6), together with the constraint equation, form a system of four equations from which four unknown variables x, y, z and λ can be determined. It is interesting to notice that, given our only interest in the extremum coordinates, we are not really interested in determining λ; this is why the technique is also known as Lagrange's method of *undetermined multipliers*. It needs also to be mentioned that the method fails if all partial derivatives of ϕ are zero at the extremum; in such an unfortunate circumstance condition d$F = 0$ is equivalent to d$f = 0$, making the introduction of multiplier λ a useless device.

Let us now summarize Lagrange's method of undetermined multipliers by extending it to n variables. To find extrema for an n-variables function, $f(x_1, x_2, ..., x_n)$, subject to $m < n$ constraints

$$\begin{cases} \phi_1(x_1, x_2, ..., x_n) = 0 \\ \phi_2(x_1, x_2, ..., x_n) = 0 \\ \quad\vdots \\ \phi_m(x_1, x_2, ..., x_n) = 0 \end{cases} \tag{7}$$

we introduce m multipliers λ_1, λ_2, ..., λ_m and form the following function:

$$\begin{aligned} F(x_1, x_2, ..., x_n) \equiv\ & f(x_1, x_2, ..., x_n) + \lambda_1\phi_1(x_1, x_2, ..., x_n) \\ & + \lambda_2\phi_2(x_1, x_2, ..., x_n) + \cdots + \lambda_m\phi_m(x_1, x_2, ..., x_n) \end{aligned} \tag{8}$$

James Foadi - Oxford 2011

By differentiating F and setting the obtained expression equal to zero, we get

$$\begin{aligned} \mathrm{d}F = & \left(\frac{\partial f}{\partial x_1} + \lambda_1 \frac{\partial \phi_1}{\partial x_1} + \cdots + \lambda_m \frac{\partial \phi_m}{\partial x_1}\right) \mathrm{d}x_1 + \\ & \vdots \\ & + \left(\frac{\partial f}{\partial x_n} + \lambda_1 \frac{\partial \phi_1}{\partial x_n} + \cdots + \lambda_m \frac{\partial \phi_m}{\partial x_n}\right) \mathrm{d}x_n = 0 \end{aligned} \tag{9}$$

We know that not all $\mathrm{d}x_j$'s are independent; we then fix the m Lagrange multipliers to satisfy the m equations,

$$\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial \phi_1}{\partial x_j} + \cdots + \lambda_m \frac{\partial \phi_m}{\partial x_j} = 0 \ , \ j = n-m+1, n-m+2, ..., n \tag{10}$$

Given the arbitrariness and independence of the remaining $n - m$ variables, condition (9) will yield, furthermore:

$$\frac{\partial f}{\partial x_j} + \lambda_1 \frac{\partial \phi_1}{\partial x_j} + \cdots + \lambda_m \frac{\partial \phi_m}{\partial x_j} = 0 \ , \ j = 1, 2, ..., n-m \tag{11}$$

Equations (10) and (11), together with equation (7), form a system of $n + m$ equations to find the n variables x_1, x_2, ..., x_n, and the m multipliers λ_1, λ_2, ..., λ_m. Solutions to this system will give coordinates for the constrained extremum.

Let us spend a few words on the nature of all found extrema, to conclude. We can say that in general this is made clear either by the kind of problem under scrutiny, or by direct substitution of such extrema in the given function. A complete procedure would be to investigate the function values in a small neighbourhood of every individual extremum.

Few problems are described in the next section, to show how the method is customarily applied in physics and mathematics.

3 Lagrange's method of undetermined multipliers (applications)

3.1 Example

Minimise $f(x, y) = x^2 + y^2$ subject to a constraint represented by equation $y + x^2 = 1$.

It is quite clear that we are looking for an extremum of $f(x, y)$, where the constraint can be re-written as:

$$\phi(x, y) \equiv y + x^2 - 1 = 0$$

$f(x, y)$ is a function depending on two variables only, so it is possible to visualise this problem geometrically (see Figure 2). The level curves for $f(x, y)$ are circles of increasing radius, centred in $(0, 0)$. The constraint is represented by the parabola $y = -x^2 + 1$. Each level curve can have one, two, three or four intersections with the parabola. Such intersections are potentially valid solution of our problem. We can immediately anticipate, just by looking at Figure 2, that it only makes sense to look for a minimum, because there is no upper limit to the value of $f(x, y)$ under the given constraint. It is also visually easy to predict that two minima will be found at points where one of the level curves is twice tangent to the parabola. Let us now proceed with the application of Lagrange's method.
First, a Lagrange multiplier λ is introduced and a new function $F = f + \lambda\phi$ formed:

$$F(x, y) = x^2 + y^2 + \lambda(y + x^2 - 1)$$

James Foadi - Oxford 2011

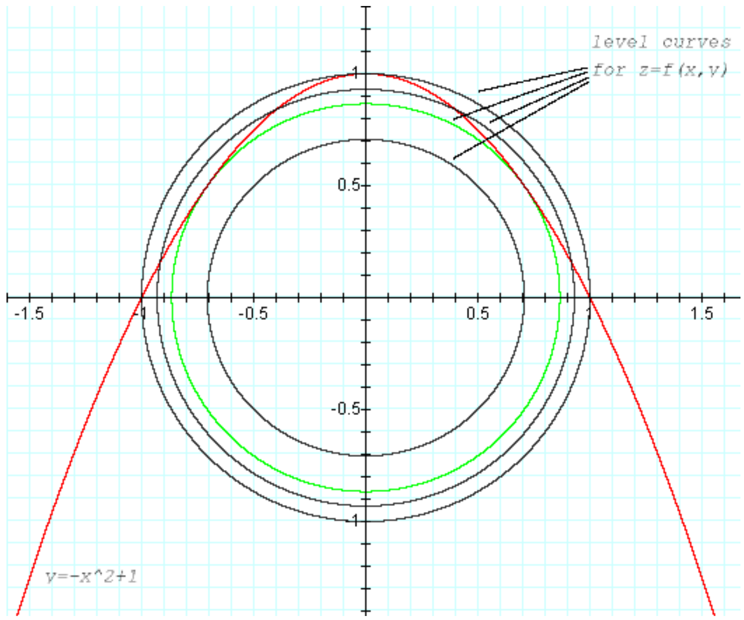

Figure 2: 2D visualization of $f(x,y) = x^2 + y^2$ and constraint $y = -x^2 + 1$.

Then we set $\partial F/\partial x$ and $\partial F/\partial y$ equal to zero and, jointly with the constraint equation, form the following system:

$$\begin{cases} 2x + 2\lambda x & = 0 \\ 2y + \lambda & = 0 \\ y + x^2 - 1 & = 0 \end{cases}$$

whose solutions are:

$$\begin{cases} x = 0 \\ y = 1 \\ \lambda = -2 \end{cases} , \quad \begin{cases} x = -\sqrt{2}/2 \\ y = 1/2 \\ \lambda = -1 \end{cases} , \quad \begin{cases} x = \sqrt{2}/2 \\ y = 1/2 \\ \lambda = -1 \end{cases}$$

Finally, computing function $f(x,y)$ at these locations:

$$f(0,1) = 1 \ , \ f(-\sqrt{2}/2, 1/2) = 3/4 \ , \ f(\sqrt{2}/2, 1/2) = 3/4$$

It is then clear that the minima we were looking for occur at $(\pm\sqrt{2}/2, 1/2)$, and $f(x,y)$ equals $3/4$ at these locations.

3.2 Example

Find the largest-area isosceles triangle inscribed in the ellipse $x^2/a^2 + y^2/b^2 = 1$, with its basis perpendicular to the x axis.

The geometry of this problem is sketched in Figure 3. If we indicate with $(-a, 0)$, (x, y) and $(x, -y)$ the three vertices of the triangle, its area will be given by $f(x,y) = (x+a)y$. We need to maximise such quantity, under the constraint that (x,y) belongs to the ellipse, $x^2/a^2 + y^2/b^2 = 1$. Once a Lagrange multiplier has been introduced and function $F(x,y)$ defined as:

$$F(x,y) = (x+a)y + \lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1\right)$$

James Foadi - Oxford 2011

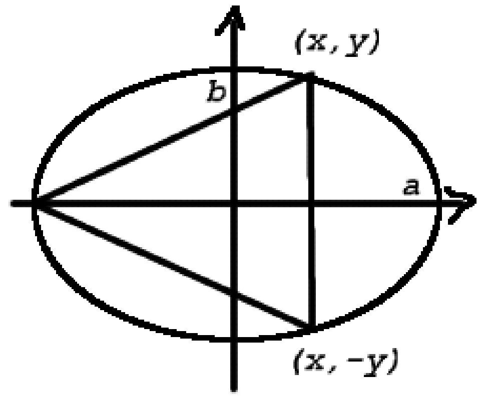

Figure 3: Inscribed largest-area triangle.

we will only need to set F's first derivatives to zero and solve system:

$$\begin{cases} y + 2\lambda x/a^2 & = & 0 \\ x + a + 2\lambda y/b^2 & = & 0 \\ x^2/a^2 + y^2/b^2 & = & 1 \end{cases}$$

The above system has three solutions, $(-a, 0)$, $(a/2, \sqrt{3}b/2)$ and $(a/2, -\sqrt{3}b/2)$. The first solution clearly corresponds to the triangle degenerating into a point, and can easily be discarded. The last two solutions are symmetrical about the x axis. They are the solutions we were looking for, corresponding to a maximum area of $f(a/2, \pm\sqrt{3}b/2) = 3\sqrt{3}ab/4$.

3.3 Example

Find the maximum and minimum values of:

$$f(x, y, z) = x - 2y + 5z$$

on the sphere $x^2 + y^2 + z^2 = 30$.

Although there are now three variables, there is still only one constraint, represented by:

$$\phi(x, y, z) = x^2 + y^2 + z^2 - 30 = 0$$

The extrema are readily found by forming:

$$F(x, y, z) = f(x, y, z) + \lambda\phi(x, y, z) = x - 2y + 5z + \lambda(x^2 + y^2 + z^2 - 30)$$

From this expression we compute F's first derivatives and equal them to zero; the derived expressions, together with $\phi = 0$, form the following system:

$$\begin{cases} 1 + 2\lambda x & = & 0 \\ -2 + 2\lambda y & = & 0 \\ 5 + \lambda z & = & 0 \\ x^2 + y^2 + z^2 & = & 30 \end{cases}$$

It has two solutions, $(1, -2, 5)$ and $(-1, 2, -5)$. There are, thus, two extrema; by substituting their values into $f(x, y, z)$ we get 30 and -30, respectively. We can then conclude that $(1, -2, 5)$ is a maximum, while $(-1, 2, -5)$ is a minimum for $x - 2y + 5z$, under the given constraint.

James Foadi - Oxford 2011

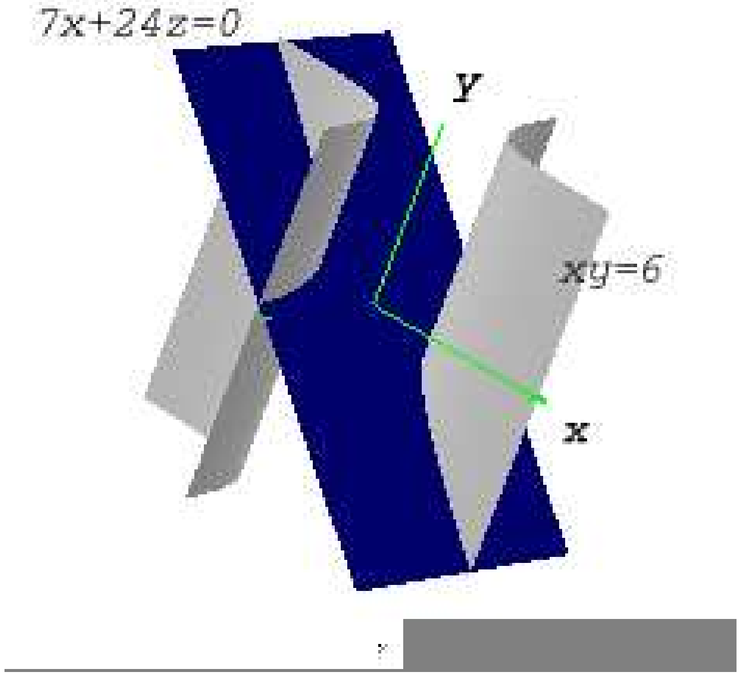

Figure 4: Geometric setting for the problem with two Lagrange's multipliers.

3.4 Example

Let us now attempt the solution of a case in which two Lagrange's multipliers need to be introduced. Find the minimum distance from the origin to the intersection of $xy = 6$ and $7x+24z = 0$.

From Figure 4 it is clear that the locus of intersection is a curve with two branches. Using the standard notation adopted in section 2, we have $f(x, y, z) = x^2 + y^2 + z^2$, because the distance to a point (x, y, z) from the origin is the quantity to minimise. Furthermore such a point has to reside on the intersection of two surfaces, i.e. it needs to satisfy both $\phi_1(x, y, z) \equiv xy - 6 = 0$ and $\phi_2(x, y, z) \equiv 7x + 24z = 0$ at the same time. Next, we introduce two Lagrange multipliers, λ_1 and λ_2, and readily define $F(x, y, z)$:

$$F(x, y, z) \equiv x^2 + y^2 + z^2 + \lambda_1(xy - 6) + \lambda_2(7x + 24z)$$

Setting F's first derivatives equal to zero, and using our constraint equations, the following 5-equations system is obtained:

$$\begin{cases} 2x + \lambda_1 y + 7\lambda_2 & = & 0 \\ 2y + \lambda_1 x & = & 0 \\ 2z + 24\lambda_2 & = & 0 \\ xy & = & 6 \\ 7x + 24z & = & 0 \end{cases}$$

from which we find two solutions, $(12/5, 5/2, -7/10)$ and $(-12/5, -5/2, 7/10)$. They both correspond to a distance from the origin equal to $5/\sqrt{2}$.

3.5 Example (from Arfken-Weber, p. 975)

Let us conclude with an example applied to physics. Consider a quantum particle of mass m in a rectangular parallelepiped box with sides a, b and c. The ground state energy of the particle is

James Foadi - Oxford 2011

given by:

$$E = \frac{h^2}{8m}\left(\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}\right)$$

We seek the shape of the box that will minimize the energy E, subject to constraint that the volume is constant,

$$V(a, b, c) = abc = K$$

In this case $E = E(a, b, c)$ is the function whose extrema we are looking for, while the constraint is given by $V(a, b, c) - K = 0$. We then introduce a Lagrange's multiplier and form the following function:

$$F(a, b, c) \equiv E(a, b, c) + \lambda[V(a, b, c) - K] = \frac{h^2}{8m}\left(\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}\right) + \lambda(abc - K)$$

From this a system is built with the usual method, resulting in the following expression:

$$\begin{cases} -h^2/(4ma^3) + \lambda bc &= 0 \\ -h^2/(4mb^3) + \lambda ac &= 0 \\ -h^2/(4mc^3) + \lambda ab &= 0 \\ abc &= K \end{cases}$$

The solution to the above system turns out to be:

$$a = b = c = K^{1/3}$$

So the ground-state energy E will be a minimum if the box is a cube.

James Foadi - Oxford 2011