

Lagrange Multipliers

The purpose of lagrange multipliers is to find the maximum value of a function $f(x, y)$ subject to a constraint $g(x, y) = k$. A way of visualizing this problem, is to find the highest elevation you reach along a mountain, supposing you are only allowed to walk a certain path.

Functions of Multiple Variables

To begin, a few definitions are needed. A function of two variables is a rule which takes 2 input values and outputs a single value. This function is said to be a mapping of an ordered pair (x, y) to a number $f(x, y)$. If the output $f(x, y) = z$, the 2 dimensional input space (x, y) can be graphed against z to create a 3 dimensional plot (x, y, z) .

A level curve of a 2 variable function is defined as the set of points (x, y) such that $f(x, y)$ is a constant value. The reason for graphing the function in a 3d space is that it gives a visual intuition for what this means. If (x, y) denotes the horizontal plane, then $f(x, y) = z$ represents the height above said plane. For constant values of $f(x, y)$, the height is therefore constant. A level set k is then the set of all points which lie on the graph (x, y, z) at height k , for constant k .

One such example of level curves being used in the real world is in topographic maps. When walking along the contour lines, your elevation will never change because that is the particular way in which the lines were defined.

Level surfaces have another use for functions of more variables. For a function with 3 inputs, $f(x, y, z)$, the entirety of the graph cannot be illustrated because it would need to be represented in 4 dimensions. However, a level set $f(x, y, z) = k$, can be represented as the set of all points in 3 dimensions that have the same-valued output.

An example of this is the temperature at any point. The value of the temperature is dependent on its position in space, thus $T = f(x, y, z)$. The temperature cannot be graphed alongside all 3 dimensions, however the set of all points at a given temperature (say, 30 °C) can be seen.

To tie back to the original problem, the constraint curve when doing lagrange multipliers $g(x, y) = k$ is an example of a level set and can be seen as a curve in (x, y) . The function $f(x, y)$ is like a mountain and we are attempting to find the highest point on the mountain while only walking along the constraint curve.

Partial Derivatives

The derivative of a single-variable function $f(x)$ can be intuitively understood as the amount of change in output $\Delta f(x)$ given a change in input Δx . In the limiting case of $x \rightarrow 0$, the derivative at the point a is said to be $\frac{d}{dx}f(a)$. In the multi-variable case, the change in $f(x, y)$ is dependent on changes in all input variables (in this case (x, y)).

We will arbitrarily resolve this problem by taking ‘partial derivatives’ with respect to each of the input variables. A partial derivative with respect to a variable x is defined as the change in output $\Delta f(x, y)$ given a change in input Δx and assuming a fixed $y = b$. Using the definition of the derivative and adding notation for the partial derivative of f with respect to x at the point (a, b) ,

$$\frac{\partial f(a, b)}{\partial x} = f_x(a, b) = \lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h}$$

Similarly, the partial derivative with respect to y is

$$\frac{\partial f(a, b)}{\partial y} = f_y(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b + h) - f(a, b)}{h}$$

For function of more than 2 variables, there will be a partial derivative with respect to each variable, where all other variables are held constant.

Chain Rule

The chain rule for a function of 1 variable $y = f(x)$ where $x = g(t)$ follows the form $\frac{dy}{dt} = \frac{dy}{dx} \frac{dx}{dt}$. In words, this means that the change in y given a change in t is equal to the change in y given a change in x multiplied by the change in x given a change in t . For functions of multiple variables, the logic is similar, however the contributions from each direction must be accounted for. This is done through addition because, for example, the x component has no effect on the y component.

Given a function $z = f(x, y)$ where $x = g(s, t)$ and $y = h(s, t)$,

$$f_t = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t}$$

Gradient Vector

Often instead of the derivative along one of the defined axes, we want to find the change in a function along a different direction. This operation will be defined as the directional derivative. If the direction is defined by a vector $\vec{u} = (a, b)$, the directional derivative must account for an a -valued change in x and a b -valued change in y . Representing this as the function $g(h) = f(x_0 + ha, y_0 + hb)$ and taking the limit as $h \rightarrow 0$ this derivative evaluated at (x_0, y_0) is represented by as,

$$g'(0) = D_{\vec{u}}f(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

Taking the same function $g(h) = f(x, y)$ where $x = x_0 + ha$ and $y = y_0 + hb$, the chain rule gives,

$$g'(h) = \frac{\partial f}{\partial x} \frac{dx}{dh} + \frac{\partial f}{\partial y} \frac{dy}{dh} = f_x(x, y) a + f_y(x, y) b$$

$$g'(0) = f_x(x_0, y_0) a + f_y(x_0, y_0) b$$

Setting the 2 pieces equal, the directional derivative of f in the direction \vec{u} is $f_x(x_0, y_0) a + f_y(x_0, y_0) b$. It is useful to rewrite this equation as the dot product between 2 vectors $(f_x, f_y) \cdot (a, b)$. Defining this first vector as the gradient of $f = \vec{\nabla} f$ and recognizing $\vec{u} = (a, b)$,

$$D_{\vec{u}}f = \vec{\nabla}f \cdot \vec{u} = |\vec{\nabla}f||\vec{u}| \cos(\theta)$$

The result of these findings is that $D_{\vec{u}}f$ has a maximum when $\cos(\theta) = 1$ or $\theta = 0$, that is when \vec{u} has the same direction as $\vec{\nabla}f$.

Lagrange Multipliers

Each of the previous sections is a necessary building block for the concept of lagrange multipliers. As before, the constraint curve when doing lagrange multipliers $g(x, y) = k$ is an example of a level set and can be seen as a curve in (x, y) . The function $f(x, y)$ is like a mountain and we are attempting to find the highest point on the mountain while only walking along the constraint curve. By drawing the level sets $f(x, y) = c$ along said mountain, both curves can be drawn in (x, y) . As an example:

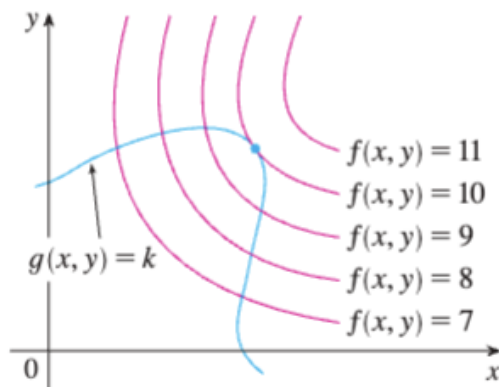


Figure 1: Level Sets

In 1d calculus, a maximum when moving in x is found when y goes from increasing to decreasing. Similarly, a maximum when moving along the curve $g(x, y) = k$ is found when $f(x, y)$ goes from increasing to decreasing. Because of this (as can be seen in [Figure 1](#)), the maximum occurs when the constraint curve touches a level set at exactly one point.

Visually, it can be seen that the 2 level surfaces are parallel to one another, and therefore their gradient vectors are parallel. Adding a factor of λ to account for the lengths of the vectors, $\vec{\nabla}f = \lambda \vec{\nabla}g$.

Making this more rigorous and showing in more dimensions, if C is the curve defined by $\vec{r}(t) = (x(t), y(t), z(t))$ the function $h(t) = f(x(t), y(t), z(t))$ represents the value of f along the curve C . If f has an extreme value at $t_0, (x_0, y_0, z_0)$, then it follows that $h'(t_0) = 0$. By the chain rule,

$$0 = f_x(x_0, y_0, z_0)x'(t_0) + f_y(x_0, y_0, z_0)y'(t_0) + f_z(x_0, y_0, z_0)z'(t_0)$$

This can be simplified down into

$$0 = \vec{\nabla} f(x_0, y_0, z_0) \cdot r'(t_0) = |\vec{\nabla} f| |r'(t_0)| \cos(\theta)$$

This finding gives that $\cos(\theta) = 0$ or that the $\vec{\nabla} f$ is normal to $r'(t_0)$. $r'(t_0)$ must be orthogonal to $\vec{\nabla} g$ by the following proof.

The function $g(x(t), y(t), z(t)) = k$, by the chain rule, gives

$$\frac{\partial g}{\partial x} \frac{dx}{dt} + \frac{\partial g}{\partial y} \frac{dy}{dt} + \frac{\partial g}{\partial z} \frac{dz}{dt} = 0$$

$$\vec{\nabla} g \cdot r'(t) = 0$$

or that the gradient of g is perpendicular to $r'(t)$. Because the gradients of f and g are orthogonal to the same vector $r'(t)$, they must be parallel, and as before, $\vec{\nabla} f = \lambda \vec{\nabla} g$.

Each input variable adds both an unknown and an equation. Because of this, using the additional constraint $g(x, y, z) = k$ will always give you enough equations to solve a system for each variable.

By solving these systems, the coordinates (x, y, z) of the maximum (or their analogs in any number of dimensions) can be found.