

# Lagrange multipliers

From Wikipedia, the free encyclopedia

In mathematical optimization, the method of **Lagrange multipliers** (named after Joseph Louis Lagrange) provides a strategy for finding the maxima and minima of a function subject to constraints.

For instance (see Figure 1 on the right), consider the optimization problem

$$\begin{aligned} &\text{maximize } f(x, y) \\ &\text{subject to } g(x, y) = c. \end{aligned}$$

We introduce a new variable ( $\lambda$ ) called a Lagrange multiplier, and study the Lagrange function defined by

$$\Lambda(x, y, \lambda) = f(x, y) + \lambda \cdot (g(x, y) - c).$$

(the  $\lambda$  term may be either added or subtracted.) If  $(x, y)$  is a maximum for the original constrained problem, then there exists a  $\lambda$  such that  $(x, y, \lambda)$  is a stationary point for the Lagrange function (stationary points are those points where the partial derivatives of  $\Lambda$  are zero). However, not all stationary points yield a solution of the original problem. Thus, the method of Lagrange multipliers yields a necessary condition for optimality in constrained problems.<sup>[1]</sup>

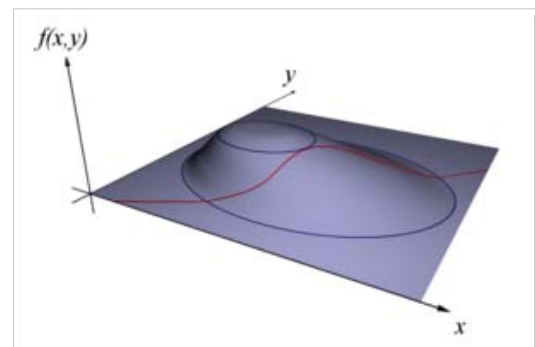


Figure 1: Find  $x$  and  $y$  to maximize  $f(x, y)$  subject to a constraint (shown in red)  $g(x, y) = c$ .

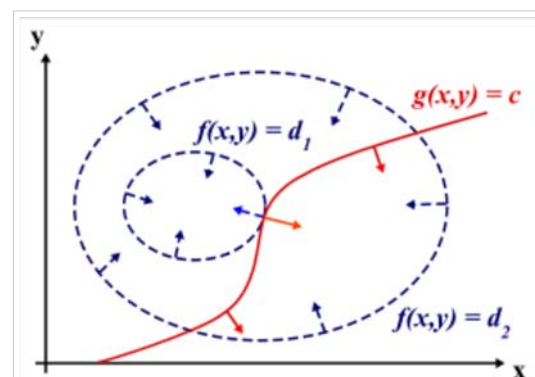


Figure 2: Contour map of Figure 1. The red line shows the constraint  $g(x, y) = c$ . The blue lines are contours of  $f(x, y)$ . The point where the red line tangentially touches a blue contour is our solution.

## Contents

- 1 Introduction
  - 1.1 Not extrema
- 2 Handling Multiple Constraints
- 3 Interpretation of the Lagrange multipliers
- 4 Examples
  - 4.1 Very simple example
  - 4.2 Simple example
  - 4.3 Example: entropy
  - 4.4 Example: numerical optimization
- 5 Applications
  - 5.1 Economics
  - 5.2 Control theory
- 6 The strong Lagrangian principle: Lagrange duality
- 7 See also
- 8 References
- 9 External links

## Introduction

Consider the two-dimensional problem introduced above:

$$\begin{aligned} &\text{maximize } f(x, y) \\ &\text{subject to } g(x, y) = c. \end{aligned}$$

We can visualize contours of  $f$  given by

$$f(x, y) = d$$

for various values of  $d$ , and the contour of  $g$  given by  $g(x, y) = c$ .

Suppose we walk along the contour line with  $g = c$ . In general the contour lines of  $f$  and  $g$  may be distinct, so following the contour line for  $g = c$  one could intersect with or cross the contour lines of  $f$ . This is equivalent to saying that while moving along the contour line for  $g = c$  the value of  $f$  can vary. Only when the contour line for  $g = c$  meets contour lines of  $f$  tangentially, we do not increase or decrease the value of  $f$ — that is, when the contour lines touch but do not cross.

The contour lines of  $f$  and  $g$  touch when the tangent vectors of the contour lines are parallel. Since the gradient of a function is perpendicular to the contour lines, this is the same as saying that the gradients of  $f$  and  $g$  are parallel. Thus we want points  $(x, y)$  where  $g(x, y) = c$  and

$$\nabla_{x,y} f = -\lambda \nabla_{x,y} g,$$

where

$$\nabla_{x,y} f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

and

$$\nabla_{x,y} g = \left( \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \right)$$

are the respective gradients. The constant  $\lambda$  is required because although the two gradient vectors are parallel, the magnitudes of the gradient vectors are generally not equal.

To incorporate these conditions into one equation, we introduce an auxiliary function

$$\Lambda(x, y, \lambda) = f(x, y) + \lambda \cdot (g(x, y) - c),$$

and solve

$$\nabla_{x,y,\lambda} \Lambda(x, y, \lambda) = 0.$$

This is the method of Lagrange multipliers. Note that  $\nabla_{\lambda} \Lambda(x, y, \lambda) = 0$  implies  $g(x, y) = c$ .

### Not extrema

The solutions are the *critical points* of the Lagrangian  $\Lambda$ ; they are not necessarily *extrema* of  $\Lambda$ . In fact, the function  $\Lambda$  is unbounded: given a point  $(x, y)$  that does not lie on the constraint, letting  $\lambda \rightarrow \pm\infty$  makes  $\Lambda$  arbitrarily large or small.

One may reformulate the Lagrangian as a Hamiltonian, in which case the solutions are local minima for the Hamiltonian. This is done in optimal control theory, in the form of Pontryagin's minimum principle.

The fact that solutions of the Lagrangian are not extrema also poses difficulties for numerical optimization. This can be addressed by computing the *magnitude* of the gradient, as the zeros of the magnitude are necessarily local minima, and is illustrated in the numerical optimization example.

## Handling Multiple Constraints

The method of *Lagrange multipliers* can also accommodate multiple constraints. To see how this is done, we need to reexamine the problem in a slightly different manner because the concept of “crossing” discussed above becomes rapidly unclear when we consider the types of constraints that are created when we have more than one constraint acting together.

As an example, consider a paraboloid with a constraint that is a single point (as might be created if we had 2 line constraints that intersect). The level set (i.e. contour line) clearly appears to “cross” that point and its gradient is clearly not parallel to the gradients of either of the two line constraints. Yet, it is obviously a maximum \*and\* a minimum because there is only one point on the paraboloid that meets the constraint.

While this example seems a bit odd, it is easy to understand and is representative of the sort of “effective” constraint that appears quite often when we deal with multiple constraints intersecting. Thus, we take a slightly different approach below to explain and derive the Lagrange Multipliers method with any number of constraints.

Throughout this section, the independent variables will be denoted by  $x_1, x_2, \dots, x_N$  and, as a group, we will denote them as  $p = (x_1, x_2, \dots, x_N)$ . Also, the function being analyzed will be denoted by  $f(p)$  and the constraints will be represented by the equations  $g_1(p) = 0, g_2(p) = 0, \dots, g_M(p) = 0$ .

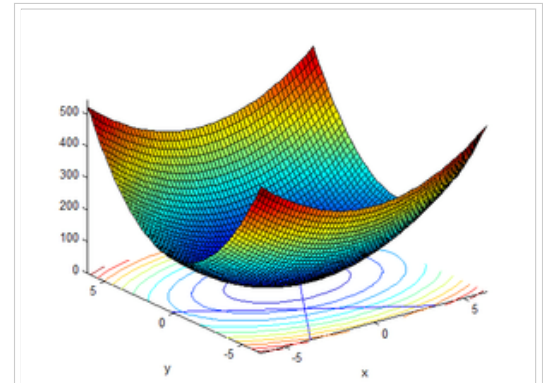
The basic idea remains essentially the same: if we consider only the points that satisfy the constraints (i.e. are *in* the constraints), then a point  $(p, f(p))$  is a stationary point (i.e. a point in a “flat” region) of  $f$  if and only if the constraints at that point do not allow movement in a direction where  $f$  changes value. It is intuitive that this is true because if the constraints allowed us to travel from this point to a (infinitesimally) near point with a *different* value, then we would *not* be in a “flat” region (i.e. a stationary point).

Once we have located the stationary points, we need to do further tests to see if we have found a minimum, a maximum or just a stationary point that is neither.

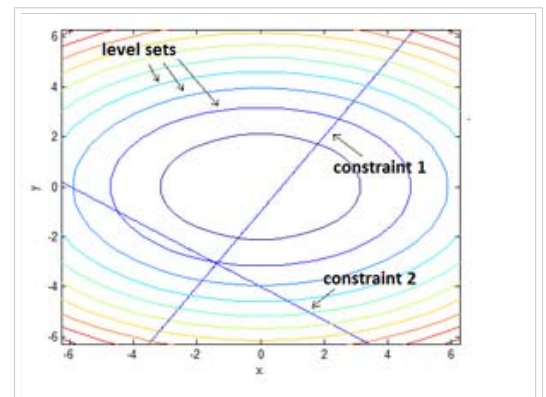
We start by considering the level set of  $f$  at  $(p, f(p))$ . The set of vectors  $\{v_L\}$  containing the directions in which we can move and still remain in the same level set are the directions where the value of  $f$  does not change (i.e. the change equals zero). Thus, for every vector  $v$  in  $\{v_L\}$ , the following relation must hold:

$$\Delta f = \frac{df}{dx_1}v_{x_1} + \frac{df}{dx_2}v_{x_2} + \dots + \frac{df}{dx_N}v_{x_N} = 0$$

Where the notation  $v_{x_K}$  above means the  $x_K$ -component of the vector  $v$ . The equation above can be rewritten in a more compact geometric form that helps our intuition:



A paraboloid, some of its level sets (aka contour lines) and 2 line constraints.



Zooming in on the levels sets and constraints, we see that the two constraint lines intersect to form a “joint” constraint that is a point. Since there is only one point to analyze, the corresponding point on the paraboloid is automatically a minimum and maximum. Yet the simplified reasoning presented in sections above seems to fail because the level set definitely appears to “cross” the point and at the same time its gradient is not parallel to the gradients of either constraint. This shows we must refine our explanation of the method to handle the kinds of constraints that are formed when we have more than one constraint acting at once.

$$\underbrace{\begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \\ \vdots \\ \frac{df}{dx_N} \end{bmatrix}}_{\nabla f} \cdot \underbrace{\begin{bmatrix} v_{x_1} \\ v_{x_2} \\ \vdots \\ v_{x_N} \end{bmatrix}}_v = 0 \quad \Rightarrow \quad \nabla f \cdot v = 0$$

This makes it clear that if we are at  $p$ , then *all* directions from this point that do *not* change the value of  $f$  must be *perpendicular* to  $\nabla f(p)$  (the gradient of  $f$  at  $p$ ).

Now let us consider the effect of the constraints. Each constraint limits the directions that we can move from a particular point and still satisfy the constraint. We can use the same procedure, to look for the set of vectors  $\{v_C\}$  containing the directions in which we can move and still satisfy the constraint. As above, for every vector  $v$  in  $\{v_C\}$ , the following relation must hold:

$$\Delta g = \frac{dg}{dx_1}v_{x_1} + \frac{dg}{dx_2}v_{x_2} + \cdots + \frac{dg}{dx_N}v_{x_N} = 0 \quad \Rightarrow \quad \nabla g \cdot v = 0$$

From this, we see that at point  $p$ , all directions from this point that will still satisfy this constraint must be perpendicular to  $\nabla g(p)$ .

Now we are ready to refine our idea further and complete the method: *a point on  $f$  is a constrained stationary point if and only if the direction that changes  $f$  violates at least one of the constraints.* (We can see that this is true because if a direction that changes  $f$  did *not* violate any constraints, then there would a “legal” point nearby with a higher or lower value for  $f$  and the current point would then not be a stationary point.)

### Single Constraint Revisited

For a *single constraint*, we use the statement above to say that at stationary points the direction that changes  $f$  is in the same direction that violates the constraint. To determine if two vectors are in the same direction, we note that if two vectors start from the same point and are “in the same direction”, then one vector can always “reach” the other by changing its length and/or flipping to point the opposite way along the same direction line. In this way, we can succinctly state that two vectors point in the same direction if and only if one of them can be multiplied by some real number such that they become equal to the other. So, for our purposes, we require that:

$$\nabla f(p) = \lambda \nabla g(p) \quad \Rightarrow \quad \nabla f(p) - \lambda \nabla g(p) = 0$$

If we now add another simultaneous equation to guarantee that we only perform this test when we are at a point that satisfies the constraint, we end up with 2 simultaneous equations that when solved, identify all constrained stationary points:

$$\begin{array}{ll} g(p) = 0 & : \text{ means point satisfies constraint} \\ \nabla f(p) - \lambda \nabla g(p) = 0 & : \text{ means point is a stationary point} \end{array}$$

Note that the above is a succinct way of writing the equations. Fully expanded, there are  $N + 1$  simultaneous equations that need

to be solved for the  $N + 1$  variables  $x_1, x_2, \dots, x_N$  and  $\lambda$ :

$$g(x_1, x_2, \dots, x_N) = 0$$

$$\frac{df}{dx_1}(x_1, x_2, \dots, x_N) - \lambda \frac{dg}{dx_1}(x_1, x_2, \dots, x_N) = 0$$

$$\frac{df}{dx_2}(x_1, x_2, \dots, x_N) - \lambda \frac{dg}{dx_2}(x_1, x_2, \dots, x_N) = 0$$

⋮

$$\frac{df}{dx_N}(x_1, x_2, \dots, x_N) - \lambda \frac{dg}{dx_N}(x_1, x_2, \dots, x_N) = 0$$

### Multiple Constraints

For more than one constraint, the same reasoning applies. If there is more than one constraint active together, each constraint contributes a direction that will violate it. Together, these “violation directions” form a “violation space”, where infinitesimal movement in any direction within the space will violate one or more constraints. Thus, to satisfy multiple constraints we can state (using this new terminology) that at the stationary points, the direction that changes  $f$  is in the “violation space” created by the constraints acting jointly.

The “violation space” created by the constraints consists of all points that can be reached by adding any combination of scaled and/or flipped versions of the individual violation direction vectors. In other words, all the points that are “reachable” when we use the individual violation directions as the basis of the space. Thus, we can succinctly state that  $v$  is in the space defined by  $b_1, b_2, \dots, b_M$  if and only if there exists a set of “multipliers”  $\lambda_1, \lambda_2, \dots, \lambda_M$  such that:

$$\sum_{k=1}^M \lambda_k b_k = v$$

Which for our purposes, translates to stating that the direction that changes  $f$  at  $p$  is in the “violation space” defined by the constraints  $g_1, g_2, \dots, g_M$  if and only if:

$$\sum_{k=1}^M \lambda_k \nabla g_k(p) = \nabla f(p) \quad \Rightarrow \quad \nabla f(p) - \sum_{k=1}^M \lambda_k \nabla g_k(p) = 0$$

As before, we now add simultaneous equation to guarantee that we only perform this test when we are at a point that satisfies every constraint, we end up with simultaneous equations that when solved, identify all constrained stationary points:

$$g_1(p) = 0$$

$$g_2(p) = 0$$

$$\vdots$$

$$g_M(p) = 0$$

: these mean point satisfies all constraints

$$\nabla f(p) - \sum_{k=1}^M \lambda_k \nabla g_k(p) = 0 \quad \text{: this means the point is a stationary point}$$

The method is complete now (from the standpoint of solving the problem of finding stationary points) but as mathematicians delight in doing, these equations can be further condensed into an even more elegant and succinct form. Lagrange must have cleverly noticed that the equations above look like partial derivatives of some larger scalar function  $L$  that takes all the  $x_1, x_2, \dots, x_N$  and all the  $\lambda_1, \lambda_2, \dots, \lambda_M$  as inputs. Next, he might then have noticed that setting every equation equal to zero is exactly what one would have to do to solve for the *unconstrained* stationary points of that larger function. Finally, he showed that a larger function  $L$  with partial derivatives that are exactly the ones we require can be constructed very simply as below:

$$L(x_1, x_2, \dots, x_N, \lambda_1, \lambda_2, \dots, \lambda_M) = f(x_1, x_2, \dots, x_N) - \sum_{k=1}^M \lambda_k g_k(x_1, x_2, \dots, x_N)$$

Solving the equation above for its *unconstrained* stationary points generates exactly the same stationary points as solving for the *constrained* stationary points of  $f$  under the constraints  $g_1, g_2, \dots, g_M$ .

In Lagrange's honor, the function above is called a *Lagrangian*, the scalars  $\lambda_1, \lambda_2, \dots, \lambda_M$  are called *Lagrange Multipliers* and this optimization method itself is called *The Method of Lagrange Multipliers*.

The method of Lagrange multipliers is generalized by the Karush–Kuhn–Tucker conditions, which can also take into account inequality constraints of the form  $h(\mathbf{x}) \leq c$ .

## Interpretation of the Lagrange multipliers

Often the Lagrange multipliers have an interpretation as some quantity of interest. To see why this might be the case, observe that:

$$\frac{\partial \Lambda}{\partial g_k} = \lambda_k.$$

So,  $\lambda_k$  is the rate of change of the quantity being optimized as a function of the constraint variable. As examples, in Lagrangian mechanics the equations of motion are derived by finding stationary points of the action, the time integral of the difference between kinetic and potential energy. Thus, the force on a particle due to a scalar potential,  $F = -\nabla V$ , can be interpreted as a Lagrange multiplier determining the change in action (transfer of potential to kinetic energy) following a variation in the particle's constrained trajectory. In economics, the optimal profit to a player is calculated subject to a constrained space of actions, where a Lagrange multiplier is the increase in the value of the objective function due to the relaxation of a given constraint (e.g. through an increase in income or bribery or other means) – the marginal cost of a constraint, called the shadow price.

In control theory this is formulated instead as costate equations.

## Examples

### Very simple example

Suppose you wish to maximize  $f(x,y) = x + y$  subject to the constraint  $x^2 + y^2 = 1$ . The constraint is the unit circle, and the level sets of  $f$  are diagonal lines (with slope -1), so one can see graphically that the maximum occurs at  $(\sqrt{2}/2, \sqrt{2}/2)$  (and the minimum occurs at  $(-\sqrt{2}/2, -\sqrt{2}/2)$ )

Formally, set  $g(x,y) - c = x^2 + y^2 - 1$ , and

$$\Lambda(x,y,\lambda) = f(x,y) + \lambda(g(x,y) - c) = x + y + \lambda(x^2 + y^2 - 1)$$

Set the derivative  $d\Lambda = 0$ , which yields the system of equations:

$$\frac{\partial \Lambda}{\partial x} = 1 + 2\lambda x = 0, \quad (\text{i})$$

$$\frac{\partial \Lambda}{\partial y} = 1 + 2\lambda y = 0, \quad (\text{ii})$$

$$\frac{\partial \Lambda}{\partial \lambda} = x^2 + y^2 - 1 = 0, \quad (\text{iii})$$

As always, the  $\partial \lambda$  equation ((iii) here) is the original constraint.

Combining the first two equations yields  $x = y$  (explicitly,  $\lambda \neq 0$ , otherwise (i) yields  $1 = 0$ , so one has  $x = -1 / (2\lambda) = y$ ).

Substituting into (iii) yields  $2x^2 = 1$ , so  $x = \pm\sqrt{2}/2$  and the stationary points are  $(\sqrt{2}/2, \sqrt{2}/2)$  and  $(-\sqrt{2}/2, -\sqrt{2}/2)$ . Evaluating the objective function  $f$  on these yields

$$f(\sqrt{2}/2, \sqrt{2}/2) = \sqrt{2} \text{ and } f(-\sqrt{2}/2, -\sqrt{2}/2) = -\sqrt{2},$$

thus the maximum is  $\sqrt{2}$ , which is attained at  $(\sqrt{2}/2, \sqrt{2}/2)$  and the minimum is  $-\sqrt{2}$ , which is attained at  $(-\sqrt{2}/2, -\sqrt{2}/2)$ .

### Simple example

Suppose you want to find the maximum values for

$$f(x, y) = x^2 y$$

with the condition that the  $x$  and  $y$  coordinates lie on the circle around the origin with radius  $\sqrt{3}$ , that is,

$$g(x, y) = x^2 + y^2 = 3.$$

As there is just a single condition, we will use only one multiplier, say  $\lambda$ .

The constraint  $g(x, y) - 3$  is identically zero on the circle of radius  $\sqrt{3}$ . So any multiple of  $g(x, y) - 3$  may be added to  $f(x, y)$  leaving  $f(x, y)$  unchanged in the region of interest (above the circle where our original constraint is satisfied). Let

$$\Lambda(x, y, \lambda) = f(x, y) + \lambda(g(x, y) - 3) = x^2 y + \lambda(x^2 + y^2 - 3).$$

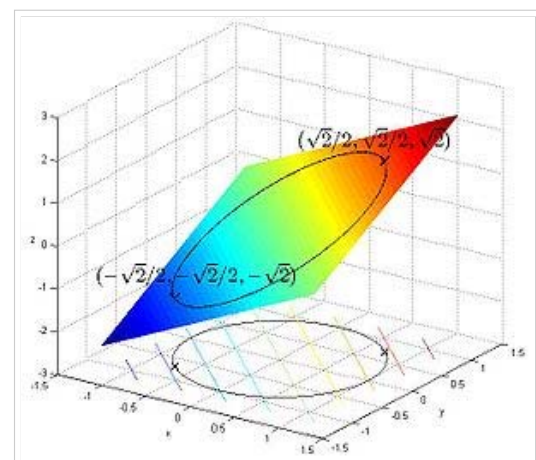


Fig. 3. Illustration of the constrained optimization problem

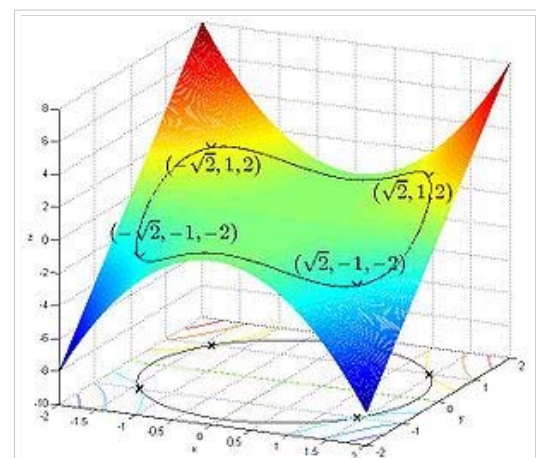


Fig. 4. Illustration of the constrained optimization problem

The critical values of  $\Lambda$  occur when its gradient is zero. The partial derivatives are

$$\frac{\partial \Lambda}{\partial x} = 2xy + 2\lambda x = 0, \quad (\text{i})$$

$$\frac{\partial \Lambda}{\partial y} = x^2 + 2\lambda y = 0, \quad (\text{ii})$$

$$\frac{\partial \Lambda}{\partial \lambda} = x^2 + y^2 - 3 = 0. \quad (\text{iii})$$

Equation (iii) is just the original constraint. Equation (i) implies  $x = 0$  or  $\lambda = -y$ . In the first case, if  $x = 0$  then we must have  $y = \pm\sqrt{3}$  by (iii) and then by (ii)  $\lambda = 0$ . In the second case, if  $\lambda = -y$  and substituting into equation (ii) we have that,

$$x^2 - 2y^2 = 0.$$

Then  $x^2 = 2y^2$ . Substituting into equation (iii) and solving for  $y$  gives this value of  $y$ :

$$y = \pm 1.$$

Thus there are six critical points:

$$(\sqrt{2}, 1); \quad (-\sqrt{2}, 1); \quad (\sqrt{2}, -1); \quad (-\sqrt{2}, -1); \quad (0, \sqrt{3}); \quad (0, -\sqrt{3}).$$

Evaluating the objective at these points, we find

$$f(\pm\sqrt{2}, 1) = 2; \quad f(\pm\sqrt{2}, -1) = -2; \quad f(0, \pm\sqrt{3}) = 0.$$

Therefore, the objective function attains a global maximum (with respect to the constraints) at  $(\pm\sqrt{2}, 1)$  and a global minimum at  $(\pm\sqrt{2}, -1)$ . The point  $(0, \sqrt{3})$  is a local minimum and  $(0, -\sqrt{3})$  is a local maximum, as may be determined by consideration of the Hessian matrix of  $\Lambda$ .

### Example: entropy

Suppose we wish to find the finite probability distribution (Without loss of generality, say on the points  $\{1, 2, \dots, n\}$ ) with maximal information entropy. Then

$$f(p_1, p_2, \dots, p_n) = - \sum_{k=1}^n p_k \log_2 p_k.$$

Of course, the sum of these probabilities equals 1, so our constraint is  $g(\mathbf{p}) = 1$  with

$$g(p_1, p_2, \dots, p_n) = \sum_{k=1}^n p_k.$$

We can use Lagrange multipliers to find the point of maximum entropy (depending on the probabilities). For all  $k$  from 1 to  $n$ , we require that

$$\frac{\partial}{\partial p_k} (f + \lambda(g - 1)) = 0,$$

which gives

$$\frac{\partial}{\partial p_k} \left( - \sum_{k=1}^n p_k \log_2 p_k + \lambda \left( \sum_{k=1}^n p_k - 1 \right) \right) = 0.$$



Carrying out the differentiation of these  $n$  equations, we get

$$-\left(\frac{1}{\ln 2} + \log_2 p_k\right) + \lambda = 0.$$

This shows that all  $p_k$  are equal (because they depend on  $\lambda$  only). By using the constraint  $\sum_k p_k = 1$ , we find

$$p_k = \frac{1}{n}.$$

Hence, the uniform distribution is the distribution with the greatest entropy, among distributions on  $n$  points.

### Example: numerical optimization

With Lagrange multipliers, the critical points occur at saddle points, rather than at local maxima (or minima). Unfortunately, many numerical optimization techniques, such as hill climbing, gradient descent, some of the quasi-Newton methods, among others, are designed to find local maxima (or minima) and not saddle points. For this reason, one must either modify the formulation to ensure that it's a minimization problem (for example, by extremizing the square of the gradient of the Lagrangian as below), or else use an optimization technique that finds stationary points (such as Newton's method without an extremum seeking line search) and not necessarily extrema.

As a simple example, consider the problem of finding the value of  $x$  that minimizes  $f(x) = x^2$ , constrained such that  $x^2 = 1$ . (This problem is somewhat pathological because there are only two values that satisfy this constraint, but it is useful for illustration purposes because the corresponding unconstrained function can be visualized in three dimensions.)

Using Lagrange multipliers, this problem can be converted into an unconstrained optimization problem:

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

The two critical points occur at saddle points where  $x = 1$  and  $x = -1$ .

In order to solve this problem with a numerical optimization technique, we must first transform this problem such that the critical points occur at local minima. This is done by computing the magnitude of the gradient of the unconstrained optimization problem.

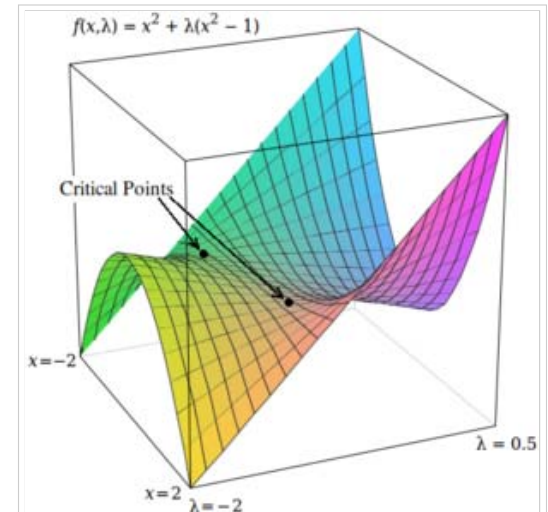
First, we compute the partial derivative of the unconstrained problem with respect to each variable:

$$\frac{\partial \Lambda}{\partial x} = 2x + 2x\lambda$$

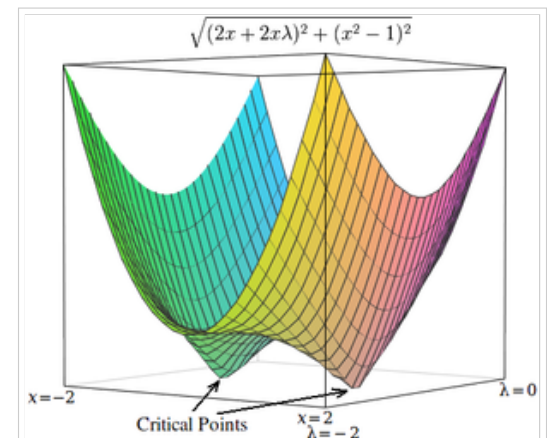
$$\frac{\partial \Lambda}{\partial \lambda} = x^2 - 1$$

If the target function is not differentiable, the differential with respect to each variable can be measured empirically:

$$\frac{\partial \Lambda}{\partial x} \approx \frac{\Lambda(x + \epsilon, \lambda) - \Lambda(x, \lambda)}{\epsilon}$$



Lagrange multipliers cause the critical points to occur at saddle points.



The magnitude of the gradient can be used to force the critical points to occur at local minima.

$$\frac{\partial \Lambda}{\partial \lambda} \approx \frac{\Lambda(x, \lambda + \epsilon) - \Lambda(x, \lambda)}{\epsilon}$$

where  $\epsilon$  is a small value.

Next, we compute the magnitude of the gradient, which is the square root of the sum of the squares of the partial derivatives:

$$h(x, \lambda) = \sqrt{(2x + 2x\lambda)^2 + (x^2 - 1)^2} \approx \sqrt{\left(\frac{\Lambda(x + \epsilon, \lambda) - \Lambda(x, \lambda)}{\epsilon}\right)^2 + \left(\frac{\Lambda(x, \lambda + \epsilon) - \Lambda(x, \lambda)}{\epsilon}\right)^2}$$

Alternatively, one may use the magnitude squared, which is the sum of the squares of the partials, without taking a square root – this has the advantage of being smooth if the partials are, while the square root may not be differentiable at the zeros.

The critical points of  $h$  occur at  $x = 1$  and  $x = -1$ , just as in  $\Lambda$ . Unlike the critical points in  $\Lambda$ , however, the critical points in  $h$  occur at local minima, so numerical optimization techniques can be used to find them.

## Applications

### Economics

Constrained optimization plays a central role in economics. For example, the choice problem for a consumer is represented as one of maximizing a utility function subject to a budget constraint. The Lagrange multiplier has an economic interpretation as the shadow price associated with the constraint, in this example the marginal utility of income.

### Control theory

In optimal control theory, the Lagrange multipliers are interpreted as costate variables, and Lagrange multipliers are reformulated as the minimization of the Hamiltonian, in Pontryagin's minimum principle.

## The strong Lagrangian principle: Lagrange duality

Given a convex optimization problem in standard form

$$\begin{aligned} & \text{minimize } f_0(x) \\ & \text{subject to } f_i(x) \leq 0, \quad i \in \{1, \dots, m\} \\ & \quad \quad h_i(x) = 0, \quad i \in \{1, \dots, p\} \end{aligned}$$

with the domain  $\mathcal{D} \subset \mathbb{R}^n$  having non-empty interior, the **Lagrangian function**  $\Lambda : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$  is defined as

$$\Lambda(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x).$$

The vectors  $\lambda$  and  $\nu$  are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem. The **Lagrange dual function**  $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$  is defined as

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \Lambda(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left( f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right).$$

The dual function  $g$  is concave, even when the initial problem is not convex. The dual function yields lower bounds on the optimal value  $p^*$  of the initial problem; for any  $\lambda \geq 0$  and any  $\nu$  we have  $g(\lambda, \nu) \leq p^*$ . If a constraint qualification such as Slater's condition holds and the original problem is convex, then we have strong duality, i.e.  $d^* = \max_{\lambda \geq 0, \nu} g(\lambda, \nu) = \inf f_0 = p^*$ .

## See also

- Karush–Kuhn–Tucker conditions: generalization of the method of Lagrange multipliers.
- Lagrange multipliers on Banach spaces: another generalization of the method of Lagrange multipliers.

## References

- ↑ Vapnyarskii, I.B. (2001), "Lagrange multipliers" (<http://eom.springer.de/L/1057190.htm>) , in Hazewinkel, Michiel, *Encyclopaedia of Mathematics*, Springer, ISBN 978-1556080104, <http://eom.springer.de/L/1057190.htm>.

## External links

### Exposition

- Conceptual introduction (<http://www.slimy.com/~steuard/teaching/tutorials/Lagrange.html>) (plus a brief discussion of Lagrange multipliers in the calculus of variations as used in physics)

### For additional text and interactive applets

- Simple explanation with an example of governments using taxes as Lagrange multipliers ([http://www.umiacs.umd.edu/~resnik/ling848\\_fa2004/lagrange.html](http://www.umiacs.umd.edu/~resnik/ling848_fa2004/lagrange.html))
- Applet (<http://ocw.mit.edu/ans7870/18/18.02/f07/tools/LagrangeMultipliersTwoVariables.html>)
- Tutorial and applet (<http://www.math.gatech.edu/~carlen/2507/notes/lagMultipliers.html>)
- Video Lecture of Lagrange Multipliers ([http://midnighttutor.com/Lagrange\\_multiplier.html](http://midnighttutor.com/Lagrange_multiplier.html))
- Slides accompanying Bertsekas's nonlinear optimization text ([http://www.athenasc.com/NLP\\_Slides.pdf](http://www.athenasc.com/NLP_Slides.pdf)) , with details on Lagrange multipliers (lectures 11 and 12)
- <http://eom.springer.de/L/1057190.htm>

Retrieved from "[http://en.wikipedia.org/wiki/Lagrange\\_multipliers](http://en.wikipedia.org/wiki/Lagrange_multipliers)"

Categories: Multivariable calculus | Mathematical optimization | Mathematical and quantitative methods (economics)

---

- This page was last modified on 2 November 2010 at 13:33.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.  
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.