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 maximum/minimum of a function subject to constraints.

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

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

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

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

If (x,y) is a maximum for the original constrained problem, then there exists a λ such that (x,y,λ) is a stationary point for the Lagrange function (stationary points are those points where the partial derivatives of Λ 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.[1]
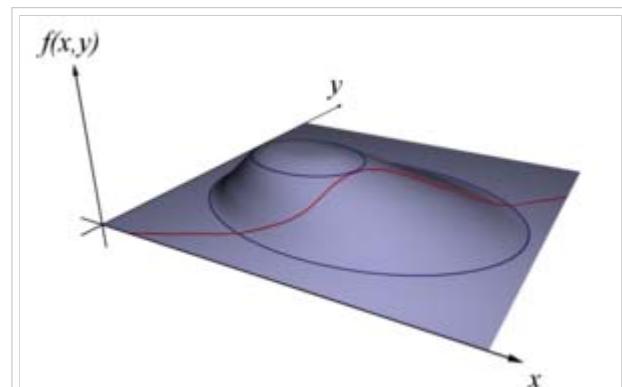

Figure 1: Find x and y to maximize $f(x,y)$ subject to a constraint (shown in red) $g(x,y) = c$.
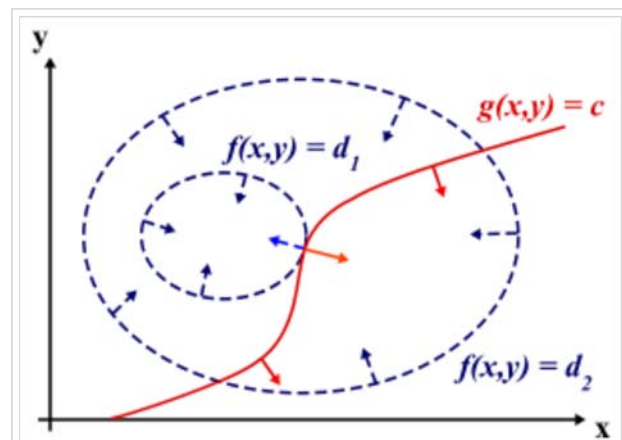

Figure 2: Contour map of Figure 1. The red line shows the constraint $g(x,y) = c$. The blue lines are contours of $f(x,y)$. The point where the red line tangentially touches a blue line is our solution.

Contents

Introduction

Consider the two-dimensional problem introduced above:

maximize $f(x, y)$
subject to $g(x, y) = c.$

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 traversing the contour line for $g = c$ 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 $g = c$ intersects 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. A familiar example can be obtained from weather maps, with their contour lines for temperature and pressure: the constrained extrema will occur where the superposed maps show touching lines (isopleths).

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

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

is the gradient. The constant λ is required because, even though the directions of both gradient vectors are equal, the magnitudes of the gradient vectors are generally not equal.

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

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

and solve

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

This is the method of Lagrange multipliers.

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

A more general formulation: The weak Lagrangian principle

Denote the objective function by $f(\mathbf{x})$ and let the constraints be given by $g_k(\mathbf{x}) = 0$. The domain of f should be an open set containing all points satisfying the constraints. Furthermore, f and the g_k must have continuous first partial derivatives and the gradients of the g_k must not be zero on the domain.[2] Now, define the Lagrangian, Λ, as

$$\Lambda(\mathbf{x},\boldsymbol{\lambda}) = f + \sum_k \lambda_k g_k.$$

k is an index for variables and functions associated with a particular constraint, k.

$\boldsymbol{\lambda}$ without a subscript indicates the vector with elements λ_k, which are taken to be independent variables.

Observe that both the optimization criteria and constraints $g_k(x)$ are compactly encoded as stationary points of the Lagrangian:

$$\nabla_{\mathbf{x}}\Lambda = \mathbf{0} \text{ if and only if } \nabla_{\mathbf{x}}f = -\sum_k \lambda_k \nabla_{\mathbf{x}}g_k,$$

$\nabla_{\mathbf{x}}$ means to take the gradient only with respect to each element in the vector $\mathbf{x}$, instead of all variables.

and

$$\nabla_{\lambda}\Lambda = \mathbf{0} \text{ implies } g_k = 0.$$

Collectively, the stationary points of the Lagrangian,

$$\nabla\Lambda = \mathbf{0},$$

give a number of unique equations totaling the length of $\mathbf{x}$ plus the length of $\boldsymbol{\lambda}$.

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, λ_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 value of relaxing a given constraint (e.g. through bribery or other means).

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 $\left(\sqrt{2}/2, \sqrt{2}/2\right)$ (and the minimum occurs at $\left(-\sqrt{2}/2, -\sqrt{2}/2\right)$)

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

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

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

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

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

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

Fig. 3. Illustration of the constrained optimization problem.

As always, the $\partial \lambda$ equation 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 $\left(\sqrt{2}/2, \sqrt{2}/2\right)$ and $\left(-\sqrt{2}/2, -\sqrt{2}/2\right)$. 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 $\left(\sqrt{2}/2, \sqrt{2}/2\right)$ and the minimum is $-\sqrt{2}$, which is attained at $\left(-\sqrt{2}/2, -\sqrt{2}/2\right)$.
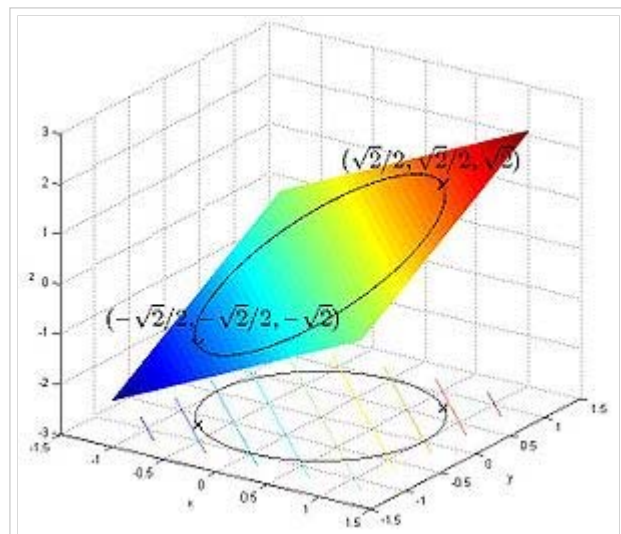
Simple example

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,

$$x^2 + y^2 = 3.$$

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

Use the constraint to define a function $g(x, y)$:

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

The function g is identically zero on the circle of radius $\sqrt{3}$.

Fig. 4. Illustration of the constrained optimization problem.

So any multiple of $g(x, y)$ 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

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

The critical values of Λ occur when its gradient is zero. The partial derivatives are

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

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

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

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

$$f(\pm\sqrt{2}, 1) = 2; \quad f(\pm\sqrt{2}, -1) = -2; \quad f(0, \pm\sqrt{3}) = 0.$$
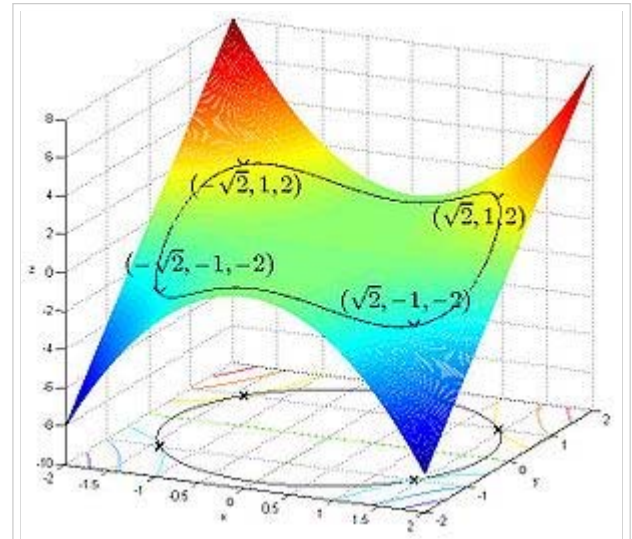
Therefore, the objective function attains a global maximum (with respect to the constraints) at $\left(\pm\sqrt{2}, 1\right)$ and a global minimum at $\left(\pm\sqrt{2}, -1\right)$. The point $\left(0, \sqrt{3}\right)$ is a local minimum and $\left(0, -\sqrt{3}\right)$ is a local maximum.

Example: entropy

Suppose we wish to find the discrete probability distribution with maximal information entropy. Then

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

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

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

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

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

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_i are equal (because they depend on λ 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.

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 case the marginal utility of income.

The strong Lagrangian principle: Lagrange duality

Given a convex optimization problem in standard form

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

with the domain $\mathcal{D} \subset \mathbb{R}^n$ having non-empty interior, the **Lagrangian function** $\Lambda : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 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 λ and ν are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem. The **Lagrange dual function** $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ 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 $\lambda \ge 0$ and any ν we have $g(\lambda, \nu) \le p^*$. If a constraint qualification such as Slater's condition holds and the original problem is convex, then we have strong duality, i.e. $d^* = \max_{\lambda \ge 0, \nu} g(\lambda, \nu) = \inf f_0 = p^*$.

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.
- Constrained optimization and Lagrange multipliers

References

1. **^** Vapnyarskii, I.B. (2001), "Lagrange multipliers", in Hazewinkel, Michiel, *Encyclopaedia of Mathematics*, Kluwer Academic Publishers, ISBN 978-1556080104.
2. **^** Gluss, David and Weisstein, Eric W., *Lagrange Multiplier* at MathWorld.

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)
- Applet (http://www-math.mit.edu/18.02/applets/LagrangeMultipliersTwoVariables.html)
- Tutorial and applet (http://www.math.gatech.edu/~carlen/2507/notes/lagMultipliers.html)
- Good Video Lecture of Lagrange Multipliers (http://midnighttutor.com/Lagrange_multiplier.html)
- Slides accompanying Bertsekas's nonlinear optimization text (http://www.athenasc.com/NLP_Slides.pdf) , with details on Lagrange multipliers (lectures 11 and 12)
- http://eom.springer.de/L/l057190.htm

Retrieved from "http://en.wikipedia.org/wiki/Lagrange_multipliers"
Categories: Mathematical economics | Multivariable calculus | Mathematical optimization
Hidden categories: Articles to be merged since December 2008 | All articles to be merged