Lagrange multiplier

From Wikipedia, the free encyclopedia

In mathematical optimization, the method of Lagrange multipliers (named after Joseph Louis Lagrange[1]) is a strategy for finding the local maxima and minima of a function subject to equality constraints.

For the case of only one constraint and only two choice variables (as exemplified in Figure 1), consider the optimization problem

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

We assume that both f and g have continuous first partial derivatives. We introduce a new variable (λ) called a Lagrange multiplier and study the Lagrange function (or Lagrangian or Lagrangian expression) defined by

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot (g(x, y) - c),$$

where the λ term may be either added or subtracted. If $f(x_0, y_0)$ is a maximum of $f(x, y)$ for the original constrained problem, then there exists λ_0 such that (x_0, y_0, λ_0) is a

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 the maximum of $f(x, y)$, since $d_1 > d_2$.

stationary point for the Lagrange function (stationary points are those points where the partial derivatives of $\mathcal{L}$ 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.[2][3][4][5][6] Sufficient conditions for a minimum or maximum also exist.

For the general case of an arbitrary number n of choice variables and an arbitrary number M of constraints, the Lagrangian takes the form

$$\mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M) = f(x_1, \ldots, x_n) - \sum_{k=1}^{M} \lambda_k g_k(x_1, \ldots, x_n),$$

where the constant term of each constraint has been subsumed in the function g_k so the kth constraint is g_k=0; again the constrained optimum of f coincides with a stationary point of $\mathcal{L}$.

Contents

Context and method

One of the most common problems in calculus is that of finding maxima or minima (in general, "extrema") of a function, but it is often difficult to find a closed form for the function being extremized. Such difficulties often arise when one wishes to maximize or minimize a function subject to fixed outside equality constraints. The method of Lagrange multipliers is a powerful tool for solving this class of problems without the need to explicitly solve the conditions and use them to eliminate extra variables.
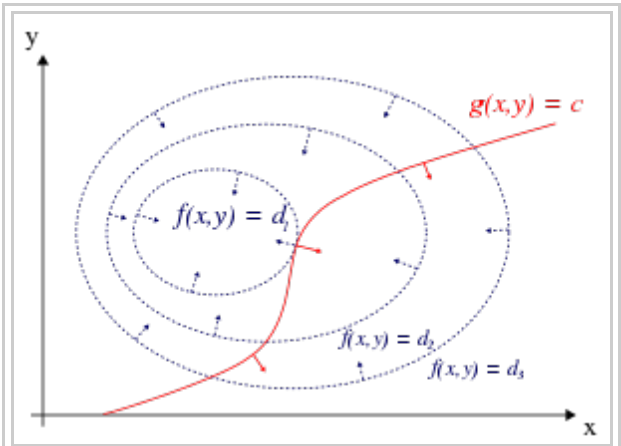
Single constraint

Consider the two-dimensional problem introduced above

> maximize $f(x, y)$
> subject to $g(x, y) = 0$.

The method of Lagrange multipliers relies on the intuition that at a maximum, $f(x, y)$ cannot be increasing in the direction of any neighboring point where $g = 0$. If it were, we could walk along $g = 0$ to get higher, meaning that the starting point wasn't actually the maximum.

We can visualize contours of f given by $f(x, y) = d$ for various values of d, and the contour of g given by $g(x, y) = 0$.

Suppose we walk along the contour line with $g = 0$. We are interested in finding points where f does not change as we walk, since these points might be maxima. There are two ways this could happen: First, we could be following a contour line of f, since by definition f does not change as we walk along its contour lines. This would mean that the contour lines of f and g are parallel here. The second possibility is that we have reached a "level" part of f, meaning that f does not change in any direction.

To check the first possibility, notice that since the gradient of a function is perpendicular to the contour lines, the contour lines of f and g are parallel if and only if the gradients of f and g are parallel. Thus we want points (x, y) where $g(x, y) = 0$ and

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

for some λ

where

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

are the respective gradients. The constant λ is required because although the two gradient vectors are parallel, the magnitudes of the gradient vectors are generally not equal. This constant is called the Lagrange multiplier. (In some conventions λ is preceded by a minus sign).

Notice that this method also solves the second possibility: if f is level, then its gradient is zero, and setting $\lambda = 0$ is a solution regardless of g.

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

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot g(x, y),$$

and solve

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

Note that this amounts to solving three equations in three unknowns. This is the method of Lagrange multipliers. Note that $\nabla_\lambda \mathcal{L}(x, y, \lambda) = 0$ implies $g(x, y) = 0$. To summarize

$$\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = 0 \iff \begin{cases} \nabla_{x,y} f(x, y) = \lambda \nabla_{x,y} g(x, y) \\ g(x, y) = 0 \end{cases}$$

The method generalizes readily to functions on n variables

$$\nabla_{x_1,\dots,x_n,\lambda} \mathcal{L}(x_1, \dots, x_n, \lambda) = 0$$

which amounts to solving $n + 1$ equations in $n + 1$ unknowns.

The constrained extrema of f are *critical points* of the Lagrangian $\mathcal{L}$, but they are not necessarily *local extrema* of $\mathcal{L}$ (see Example 2 below).

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 necessarily 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, as illustrated in the numerical optimization example.

Multiple constraints

The method of Lagrange multipliers can be extended to solve problems with multiple constraints using a similar argument. Consider a paraboloid subject to two line constraints that intersect at a single point. As the only feasible solution, this point is obviously a constrained extremum. However, the level set of f is clearly not parallel to either constraint at the intersection point (see Figure 3); instead, it is a linear combination of the two constraints' gradients. In the case of multiple constraints, that will be what we seek in general: the method of Lagrange seeks points not at which the gradient of f is multiple of any single constraint's gradient necessarily,

but in which it is a linear combination of all the constraints' gradients.

Concretely, suppose we have M constraints and are walking along the set of points satisfying $g_i(\mathbf{x}) = 0, i = 1, \ldots, M$. Every point $\mathbf{x}$ on the contour of each constraint function g has a space of allowable directions: the space of vectors perpendicular to $\nabla g(\mathbf{x})$. The set of directions that are allowed by all constraints is thus the space of directions perpendicular to all of the constraints' gradients. Denote this space of allowable moves by A and denote the span of the constraints' gradients by S. Then $A = S^\perp$, the space of vectors perpendicular to every element of S.

We are still interested in finding points where f does not change as we walk, since these points might be (constrained) extrema. We therefore seek $\mathbf{x}$ such that any allowable direction of movement away from $\mathbf{x}$ is perpendicular to $\nabla f(\mathbf{x})$ (otherwise we could increase f by moving along that allowable direction). In other words, $\nabla f(\mathbf{x}) \in A^\perp = S$. Thus there are scalars $\lambda_1, \lambda_2, \ldots, \lambda_M$ such that

Figure 2: A paraboloid constrained along two intersecting lines.

$$\nabla f(\mathbf{x}) = \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) \quad \Longleftrightarrow \quad \nabla f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) = 0.$$

These scalars are the Lagrange multipliers. We now have M of them, one for every constraint.

As before, we introduce an auxiliary function

$$\mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M) = f(x_1, \ldots, x_n) - \sum_{k=1}^{M} \lambda_k g_k(x_1, \ldots, x_n)$$

and solve

Figure 3: Contour map of Figure 2.

$$\nabla_{x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M} \mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M) = 0 \iff \begin{cases} \nabla f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) = 0 \\ g_1(\mathbf{x}) = \cdots = g_M(\mathbf{x}) = 0 \end{cases}$$

which amounts to solving $n + M$ equations in $n + M$ unknowns.

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

Modern formulation via differentiable manifolds

The problem of finding the local maxima and minima subject to equality constraints can be generalized to finding local maxima and minima on a differentiable manifold. Let M be a smooth manifold. Suppose $f : M \to \mathbb{R}$ is a differentiable function we wish to maximize. By an extension of Fermat's theorem the local maxima are only at points where the exterior derivative df vanishes.[7] In particular, supposing we have a local coordinate chart $\varphi : V \to \mathbb{R}^n$ for $V \subseteq M$, the extrema are a subset of the critical points of the function $f' = f \circ \varphi^{-1} : U \to \mathbb{R}$. However, it is difficult to compute φ^{-1} explicitly. *The entire method of Lagrange multipliers reduces to the idea of skipping that step and finding the zeros of df' directly.*

In particular, suppose the manifold M defined by a smooth level set function $g : \mathbb{R}^n \to \mathbb{R}^p$ as $M = \{x \in R^n : g(x) = c\}$, with $c \in \mathbb{R}^p$ and dg a submersion. It follows from the construction in the level set theorem that the image of $D_x \varphi^{-1}$ is $\ker D_{\varphi^{-1}(x)} g \subseteq \mathbb{R}^n$. Therefore,

$$0 = D_x f' = D_x(f \circ \varphi^{-1}) = D_{\varphi^{-1}(x)} f \circ D_x \varphi^{-1}$$

if and only if

$$\ker D_y g \subseteq \ker D_y f$$

writing y for $\varphi^{-1}(x)$.

By the first isomorphism theorem this is true if and only if there exists a linear map $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ such that $\mathcal{L} \circ D_y g = D_y f$. As a linear map, we must have that $\mathcal{L}(x) = \lambda x$ for a fixed $\lambda \in \mathbb{R}^p$. So finding a critical point of f' is equivalent to solving the system of equations

$$\lambda D_y g = D_y f$$
$$g(y) = c$$

in the variables $y \in \mathbb{R}^{n-1}$ and $\lambda \in \mathbb{R}$. This is in general a non-linear system of n equations and n unknowns.
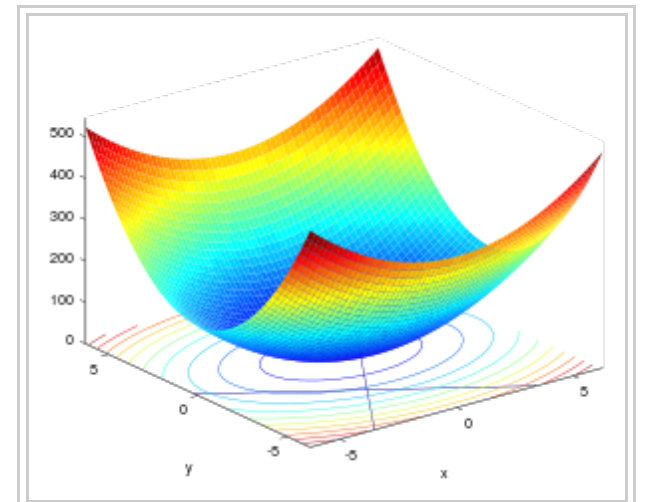
Finding local maxima of a function $f : U \to \mathbb{R}$ is done by finding all points $x \in U$ such that $D_x f = 0$ then checking whether all the eigenvalues of the Hessian $H_x f$ are negative. (Note that the maxima may not exist even if f is continuous because U is open, and also note that the conditions checked here are sufficient but not necessary for optimality.) Setting $D_x f = 0$ is a non-linear problem and in general, arbitrarily difficult. After finding the critical points, checking the eigenvalues is a linear problem and thus easy.

Interpretation of the Lagrange multipliers

Often the Lagrange multipliers have an interpretation as some quantity of interest. For example, if the Lagrangian expression is

$$\mathcal{L}(x_1, x_2, \ldots; \lambda_1, \lambda_2, \ldots) = f(x_1, x_2, \ldots) + \lambda_1(c_1 - g_1(x_1, x_2, \ldots)) + \lambda_2(c_2 - g_2(x_1, x_2, \ldots)) + \ldots$$

then

$$\frac{\partial \mathcal{L}}{\partial c_k} = \lambda_k.$$

So, λ_k is the rate of change of the quantity being optimized as a function of the constraint parameter. 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 control theory this is formulated instead as costate equations.

Moreover, by the envelope theorem the optimal value of a Lagrange multiplier has an interpretation as the marginal effect of the corresponding constraint constant upon the optimal attainable value of the original objective function: if we denote values at the optimum with an asterisk, then it can be shown that

$$\frac{\mathrm{d}f(x_1^*(c_1, c_2, \ldots), x_2^*(c_1, c_2, \ldots), \ldots)}{\mathrm{d}c_k} = \lambda_k^*.$$

For example, in economics the optimal profit to a player is calculated subject to a constrained space of actions, where a Lagrange multiplier is the change in the optimal value of the objective function (profit) due to the relaxation of a given constraint (e.g. through a change in income); in such a context $\lambda*$ is the marginal cost of the constraint, and is referred to as the shadow price.

Sufficient conditions

Sufficient conditions for a constrained local maximum or minimum can be stated in terms of a sequence of principal minors (determinants of upper-left-justified sub-matrices) of the bordered Hessian matrix of second derivatives of the Lagrangian expression.[8]

Examples

Example 1

Example 1a

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

Illustration of the constrained optimization problem

Using the method of Lagrange multipliers, we have

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

hence

$$\begin{aligned} \mathcal{L}(x, y, \lambda) &= f(x, y) + \lambda \cdot g(x, y) \\ &= x + y + \lambda(x^2 + y^2 - 1). \end{aligned}$$

Now we can calculate the gradient:

$$\begin{aligned} \nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) &= \left(\frac{\partial \mathcal{L}}{\partial x}, \frac{\partial \mathcal{L}}{\partial y}, \frac{\partial \mathcal{L}}{\partial \lambda}\right) \\ &= \left(1 + 2\lambda x, 1 + 2\lambda y, x^2 + y^2 - 1\right) \end{aligned}$$
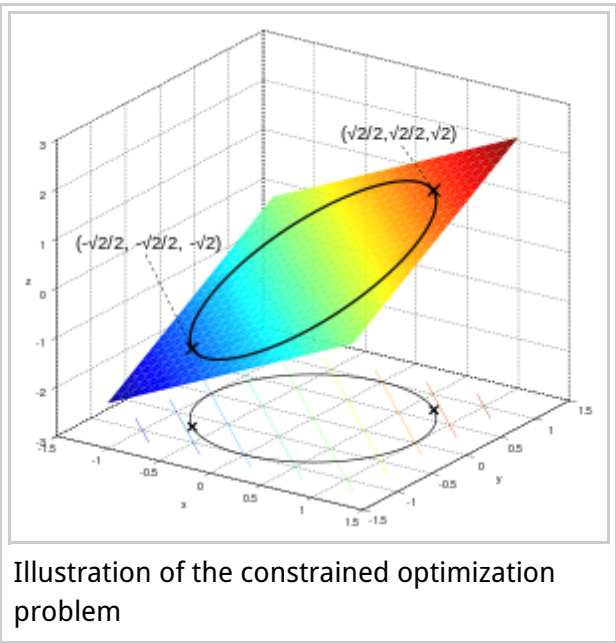
and therefore:

$$\nabla_{x,y,\lambda}\mathcal{L}(x,y,\lambda)=0 \quad \Leftrightarrow \quad \begin{cases} 1+2\lambda x=0 \\ 1+2\lambda y=0 \\ x^2+y^2-1=0 \end{cases}$$

Notice that the last equation is the original constraint.

The first two equations yield

$$x=y=-\frac{1}{2\lambda}, \qquad \lambda \neq 0.$$

By substituting into the last equation we have:

$$\frac{1}{4\lambda^2}+\frac{1}{4\lambda^2}-1=0,$$

so

$$\lambda=\pm\frac{1}{\sqrt{2}},$$

which implies that the stationary points of $\mathcal{L}$ are

$$\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2},\frac{-\sqrt{2}}{2}\right), \qquad \left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right).$$

Evaluating the objective function f at these points yields

$$f\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)=\sqrt{2}, \qquad f\left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)=-\sqrt{2}.$$

Thus the constrained maximum is $\sqrt{2}$ and the constrained minimum is $-\sqrt{2}$.

Example 1b

Now we modify the objective function of Example 1a so that we minimize $f(x,y)=(x+y)^2$ instead of $f(x,y)=x+y$, again along the circle $g(x,y)=x^2+y^2=1$. Now the level sets of f are still lines of slope -1, and the points on the circle tangent to these level sets are again $(\sqrt{2}/2, \sqrt{2}/2)$ and $(-\sqrt{2}/2, -\sqrt{2}/2)$. These tangency points are maxima of f.

On the other hand, the minima occur on the level set for f=0 (since by its construction f cannot take negative values), at $(\sqrt{2}/2, -\sqrt{2}/2)$ and $(-\sqrt{2}/2, \sqrt{2}/2)$, where the level curves of f are not tangent to the constraint. The condition that $\nabla f = \lambda \nabla g$ correctly identifies all four points as extrema; the minima are characterized in particular by $\lambda = 0$.

Example 2

In this example we will deal with some more strenuous calculations, but it is still a single constraint problem.

Suppose we want to find the maximum values of

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

with the condition that the x and y coordinates lie on the circle around the origin with radius $\sqrt{3}$, that is, subject to the constraint
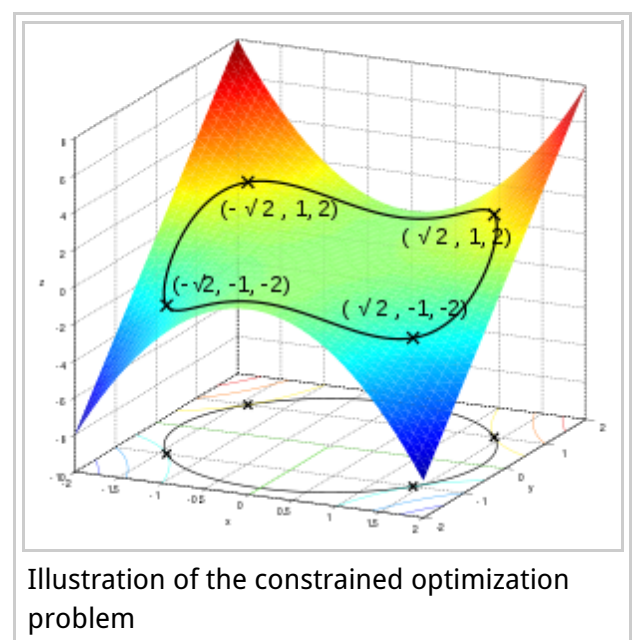
$$g(x,y)=x^2+y^2-3=0.$$

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

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

Apply the ordinary Langrange multiplier method. Let:

$$\mathcal{L}(x,y,\lambda)=f(x,y)+\lambda\cdot g(x,y)$$
$$=x^2y+\lambda(x^2+y^2-3)$$

Now we can calculate the gradient:

Illustration of the constrained optimization problem

$$\nabla_{x,y,\lambda}\mathcal{L}(x,y,\lambda) = \left(\frac{\partial\mathcal{L}}{\partial x},\frac{\partial\mathcal{L}}{\partial y},\frac{\partial\mathcal{L}}{\partial\lambda}\right)$$
$$= \left(2xy + 2\lambda x, x^2 + 2\lambda y, x^2 + y^2 - 3\right)$$

And therefore:

$$\nabla_{x,y,\lambda}\mathcal{L}(x,y,\lambda) = 0 \quad\Leftrightarrow\quad \begin{cases} 2xy + 2\lambda x = 0 \\ x^2 + 2\lambda y = 0 \\ x^2 + y^2 - 3 = 0 \end{cases} \quad\Leftrightarrow\quad \begin{cases} x(y+\lambda) = 0 & (i) \\ x^2 = -2\lambda y & (ii) \\ x^2 + y^2 = 3 & (iii) \end{cases}$$

Notice that (iii) is just the original constraint. (i) implies $x = 0$ *or* $\lambda = -y$. If $x = 0$ then $y = \pm\sqrt{3}$ by (iii) and consequently $\lambda = 0$ from (ii). If $\lambda = -y$, substituting in (ii) we get $x^2 = 2y^2$. Substituting this in (iii) and solving for y gives $y = \pm 1$. Thus there are six critical points of $\mathcal{L}$:

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

Evaluating the objective at these points, we find that

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

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

Note that while $(\sqrt{2}, 1, -1)$ is a critical point of $\mathcal{L}$, it is not a local extremum of $\mathcal{L}$. We have

$$\mathcal{L}\left(\sqrt{2} + \epsilon, 1, -1 + \delta\right) = 2 + \delta\left(\epsilon^2 + \left(2\sqrt{2}\right)\epsilon\right).$$

Given any neighborhood of $(\sqrt{2}, 1, -1)$, we can choose a small positive ϵ and a small δ of either sign to get $\mathcal{L}$ values both greater and less than 2. This can also be seen from the fact that the Hessian matrix of $\mathcal{L}$ evaluated at this point (or indeed at any of the critical points) is an indefinite matrix. Each of the critical points of $\mathcal{L}$ is a saddle point of $\mathcal{L}$.

Example 3: Entropy

Suppose we wish to find the discrete probability distribution on the points $\{p_1, p_2, \cdots, p_n\}$ with maximal information entropy. This is the same as saying that we wish to find the least structured probability distribution on the points $\{p_1, p_2, \cdots, p_n\}$. In other words, we wish to maximize the Shannon entropy equation:

$$f(p_1, p_2, \cdots, p_n) = -\sum_{j=1}^{n} p_j \log_2 p_j.$$

For this to be a probability distribution the sum of the probabilities p_i at each point x_i must equal 1, so our constraint is:

$$g(p_1, p_2, \cdots, p_n) = \sum_{j=1}^{n} p_j = 1.$$

We use Lagrange multipliers to find the point of maximum entropy, $\vec{p}\,^*$, across all discrete probability distributions $\vec{p}$ on $\{x_1, x_2, \cdots, x_n\}$. We require that:

$$\left.\frac{\partial}{\partial\vec{p}}(f + \lambda(g - 1))\right|_{\vec{p}=\vec{p}\,^*} = 0,$$

which gives a system of *n* equations, $k = 1, \cdots, n$, such that:

$$\left.\frac{\partial}{\partial p_k}\left\{-\left(\sum_{j=1}^{n} p_j \log_2 p_j\right) + \lambda\left(\sum_{j=1}^{n} p_j - 1\right)\right\}\right|_{p_k = p_k^*} = 0.$$

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 λ only). By using the constraint

$$\sum_j p_j = 1,$$

we find

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

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

Example 4: Numerical optimization

The critical points of Lagrangians occur at saddle points, rather than at local maxima (or minima).[9] 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:

$$\mathcal{L}(x, \lambda) = x^2 + \lambda(x^2 - 1).$$

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:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + 2x\lambda$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x^2 - 1.$$

If the target function is not easily differentiable, the differential with respect to each variable can be approximated as

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

where ϵ is a small value.

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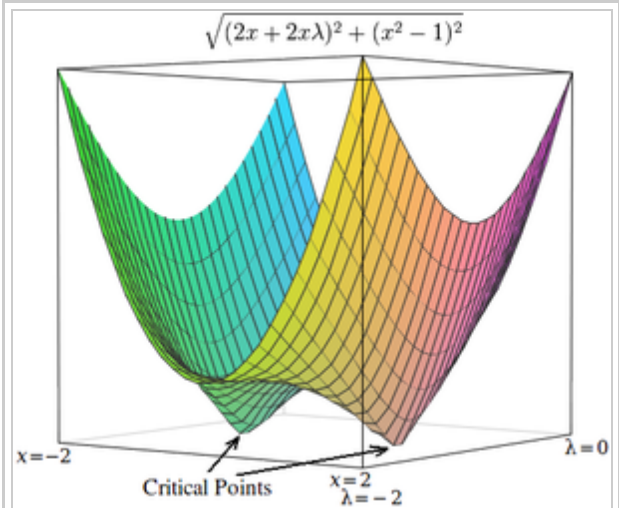
The magnitude of the gradient can be used to force the critical points to occur at local minima.

Next, we compute the magnitude of the gradient, which is the square root of the sum of the squares of the partial derivatives:
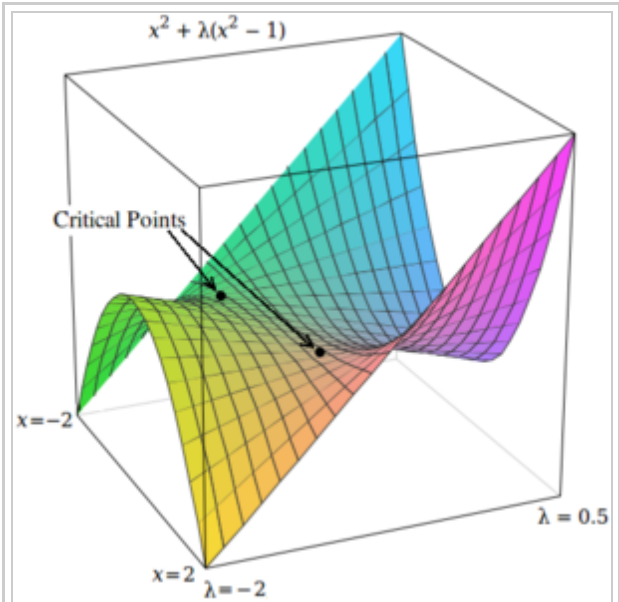
$$h(x, \lambda) = \sqrt{(2x + 2x\lambda)^2 + (x^2 - 1)^2} \approx \sqrt{\left(\frac{\mathcal{L}(x + \epsilon, \lambda) - \mathcal{L}(x, \lambda)}{\epsilon}\right)^2 + \left(\frac{\mathcal{L}(x, \lambda + \epsilon) - \mathcal{L}(x, \lambda)}{\epsilon}\right)^2}.$$

(Since magnitude is always non-negative, optimizing over the squared-magnitude is equivalent to optimizing over the magnitude. Thus, the "square root" may be omitted from these equations with no expected difference in the results of optimization.)

The critical points of h occur at $x = 1$ and $x = -1$, just as in $\mathcal{L}$. Unlike the critical points in $\mathcal{L}$, 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 constant in the constraint, in this example the marginal utility of income. Other examples include profit maximization for a firm, along with various macroeconomic applications.

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.

Nonlinear programming

The Lagrange multiplier method has several generalizations. In nonlinear programming there are several multiplier rules, *e.g.*, the Carathéodory-John Multiplier Rule and the Convex Multiplier Rule, for inequality constraints.[10]

See also

- Adjustment of observations
- Duality
- Karush–Kuhn–Tucker conditions: generalization of the method of Lagrange multipliers.
- Lagrange multipliers on Banach spaces: another generalization of the method of Lagrange multipliers.
- Lagrangian relaxation

References

1. Mécanique Analytique sect. IV, 2 vols. Paris, 1811 https://archive.org/details/mcaniqueanalyt01lagr
2. Bertsekas, Dimitri P. (1999). *Nonlinear Programming* (Second ed.). Cambridge, MA.: Athena Scientific. ISBN 1-886529-00-0.
3. Vapnyarskii, I.B. (2001), "Lagrange multipliers" (https://www.encyclopediaofmath.org/index.php?title=Lagrange_multipliers), in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4.
4. ■ Lasdon, Leon S. (1970). *Optimization theory for large systems*. Macmillan series in operations research. New York: The Macmillan Company. pp. xi+523. MR 337317 (https://www.ams.org/mathscinet-getitem?mr=337317).
   ■ Lasdon, Leon S. (2002). *Optimization theory for large systems* (reprint of the 1970 Macmillan ed.). Mineola, New York: Dover Publications, Inc. pp. xiii+523. MR 1888251 (https://www.ams.org/mathscinet-getitem?mr=1888251).
5. Hiriart-Urruty, Jean-Baptiste; Lemaréchal, Claude (1993). "XII Abstract duality for practitioners". *Convex analysis and minimization algorithms, Volume II: Advanced theory and bundle methods*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. 306. Berlin: Springer-Verlag. pp. 136–193 (and Bibliographical comments on pp. 334–335). ISBN 3-540-56852-2. MR 1295240 (https://www.ams.org/mathscinet-getitem?mr=1295240).
6. Lemaréchal, Claude (2001). "Lagrangian relaxation". In Michael Jünger and Denis Naddef. *Computational combinatorial optimization: Papers from the Spring School held in Schloß Dagstuhl, May 15–19, 2000*. Lecture Notes in Computer Science. 2241. Berlin: Springer-Verlag. pp. 112–156. ISBN 3-540-42877-1. MR 1900016 (https://www.ams.org/mathscinet-getitem?mr=1900016). doi:10.1007/3-540-45586-8_4 (https://doi.org/10.1007%2F3-540-45586-8_4).
7. Lafontaine, Jacques. *An Introduction to Differential Manifolds* (https://books.google.com/books?id=KNhJCgAAQBAJ&lpg=PA70&pg=PA70#v=onepage&q&f=false). Springer. p. 70. ISBN 9783319207353.
8. Chiang, Alpha C., *Fundamental Methods of Mathematical Economics*, McGraw-Hill, third edition, 1984: p. 386. ISBN 9757860069
9. Michael T. Heath (2005). *Scientific Computing: An Introductory Survey* (https://books.google.com/books?id=gwBrMAEACAAJ). McGraw-Hill. p. 203. ISBN 978-0-07-124489-3.
10. Pourciau, Bruce H. (1980). "Modern multiplier rules" (http://www.maa.org/programs/maa-awards/writing-awards/modern-multiplier-rules). *Amer. Math. Monthly*. 87: 433–452. doi:10.2307/2320250 (https://doi.org/10.2307%2F2320250).

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)
- Lagrange Multipliers for Quadratic Forms With Linear Constraints (http://ece.k-state.edu/people/faculty/carpenter/documents/lagrange.pdf) by Kenneth H. Carpenter

The Wikibook *Calculus optimization methods* has a page on the topic of: *Lagrange multipliers*

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)
- Lagrange Multipliers without Permanent Scarring (http://nlp.cs.berkeley.edu/tutorials/lagrange-multipliers.pdf) Explanation with focus on the intuition by Dan Klein
- Geometric Representation of Method of Lagrange Multipliers (http://demonstrations.wolfram.com/GeometricRepresentationOfMethodOfLagrangeMultipliers) Provides compelling insight in 2 dimensions that at a minimizing point, the direction of steepest descent must be perpendicular to the tangent of the constraint curve at that point. [Needs InternetExplorer/Firefox/Safari] *Mathematica* demonstration by Shashi Sathyanarayana

- Applet (http://ocw.mit.edu/ans7870/18/18.02/f07/tools/LagrangeMultipliersTwoVariables.html)
- Video Lecture of Lagrange Multipliers (http://midnighttutor.com/Lagrange_multiplier.html)
- MIT OpenCourseware Video Lecture on Lagrange Multipliers from Multivariable Calculus course (http://ocw.mit.edu/courses /mathematics/18-02-multivariable-calculus-fall-2007/video-lectures/lecture-13-lagrange-multipliers/)
- Slides accompanying Bertsekas's nonlinear optimization text (http://www.athenasc.com/NLP_Slides.pdf), with details on Lagrange multipliers (lectures 11 and 12)
- Geometric idea behind Lagrange multipliers (http://www-mtl.mit.edu/Courses/6.050/2004/unit9/wyatt.apr.7.pdf)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Lagrange_multiplier&oldid=779217088"

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