WIKIPEDIA

Lagrange multiplier

In mathematical optimization, the **method of Lagrange multipliers** is a strategy for finding the local maxima and minima of a function subject to equality constraints (i.e., subject to the condition that one or more equations have to be satisfied exactly by the chosen values of the variables).[1] It is named after the mathematician Joseph-Louis Lagrange. The basic idea is to convert a constrained problem into a form such that the derivative test of an unconstrained problem can still be applied. The relationship between the gradient of the function and gradients of the constraints rather naturally leads to a reformulation of the original problem, known as the **Lagrangian function**.[2]

The method can be summarized as follows: in order to find the maximum or minimum of a function $f(x)$ subjected to the equality constraint $g(x) = 0$, form the Lagrangian function

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

and find the stationary points of $\mathcal{L}$ considered as a function of x and the Lagrange multiplier λ.[3] The solution corresponding to the original constrained optimization is always a saddle point of the Lagrangian function,[4][5] which can be identified among the stationary points from the definiteness of the bordered Hessian matrix.[6]

The great advantage of this method is that it allows the optimization to be solved without explicit parameterization in terms of the constraints. As a result, the method of Lagrange multipliers is widely used to solve challenging constrained optimization problems. Further, the method of Lagrange multipliers is generalized by the Karush–Kuhn–Tucker conditions, which can also take into account inequality constraints of the form $h(\mathbf{x}) \leq c$ for a given constant c.

Contents

Statement

The following is known as the Lagrange multiplier theorem.[7]

Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be the objective function, $g \colon \mathbb{R}^n \to \mathbb{R}^c$ be the constraints function, both belonging to C^1 (that is, having continuous first derivatives). Let x^* be an optimal solution to the following optimization problem such that rank $Dg(x^*) = c < n$:

> maximize $f(x)$
> subject to: $g(x) = 0$

(Here $Dg(x^*)$ denotes the matrix of partial derivatives, $\left[\dfrac{\partial g_j}{\partial x_k}\right]$.)

Then there exists a unique Lagrange multiplier $\lambda^* \in \mathbb{R}^c$ such that $Df(x^*) = \lambda^{*T} Dg(x^*)$.

The Lagrange multiplier theorem states that at any local maxima (or minima) of the function evaluated under the equality constraints, if constraint qualification applies (explained below), then the gradient of the function (at that point) can be expressed as a linear combination of the gradients of the constraints (at that point), with the Lagrange multipliers acting as coefficients.[8] This is equivalent to saying that any direction perpendicular to all gradients of the constraints is also perpendicular to the gradient of the function. Or still, saying that the directional derivative of the function is 0 in every feasible direction.

Single constraint

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: $g(x, y) = 0$

(Sometimes an additive constant is shown separately rather than being included in g, in which case the constraint is written $g(x, y) = c$, as in Figure 1.) We assume that both f and g have continuous first partial derivatives. We introduce a new variable (λ) called a **Lagrange multiplier** (or

Lagrange undetermined multiplier) and study the **Lagrange function** (or **Lagrangian** or **Lagrangian expression**) defined by

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

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 and $\nabla g(x_0, y_0) \neq 0$, then there exists λ_0 such that (x_0, y_0, λ_0) is a stationary point for the Lagrange function (stationary points are those points where the first partial derivatives of $\mathcal{L}$ are zero). The assumption $\nabla g \neq 0$ is called constraint qualification. However, not all stationary points yield a solution of the original problem, as the method of Lagrange multipliers yields only a necessary condition for optimality in constrained problems.[9][10][11][12][13] Sufficient conditions for a minimum or maximum also exist, but if a particular candidate solution satisfies the
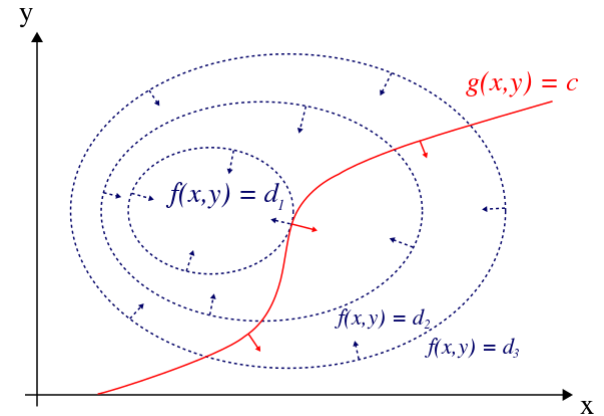

Figure 1: The red curve shows the constraint $g(x, y) = c$. The blue curves are contours of $f(x, y)$. The point where the red constraint tangentially touches a blue contour is the maximum of $f(x, y)$ along the constraint, since $d_1 > d_2$.

sufficient conditions, it is only guaranteed that that solution is the best one *locally* – that is, it is better than any permissible nearby points. The *global* optimum can be found by comparing the values of the original objective function at the points satisfying the necessary and locally sufficient conditions.

The method of Lagrange multipliers relies on the intuition that at a maximum, $f(x, y)$ cannot be increasing in the direction of any such neighboring point that also has $g = 0$. If it were, we could walk along $g = 0$ to get higher, meaning that the starting point wasn't actually the maximum. Viewed in this way, it is an exact analogue to testing if the derivative of an unconstrained function is 0, that is, we are verifying that the directional derivative is 0 in any relevant (viable) direction.

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 $g = c$. We are interested in finding points where f almost does not change as we walk, since these points might be maxima.

There are two ways this could happen:

1. We could touch a contour line of f, since by definition f does not change as we walk along its contour lines. This would mean that the tangents to the contour lines of f and g are parallel here.
2. We have reached a "level" part of f, meaning that f does not change in any direction.

To check the first possibility (we touch a contour line of f), notice that since the gradient of a function is perpendicular to the contour lines, the tangents to 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 $g(x, y) = c$ 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, that f is level: if f is level, then its gradient is zero, and setting $\lambda = 0$ is a solution regardless of $\nabla_{x,y} g$.

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

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

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 $g(x, y) = 0$, as the partial derivative of $\mathcal{L}$ with respect to λ is $-g(x, y)$, which clearly is zero if and only if $g(x, y) = 0$.

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

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

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.
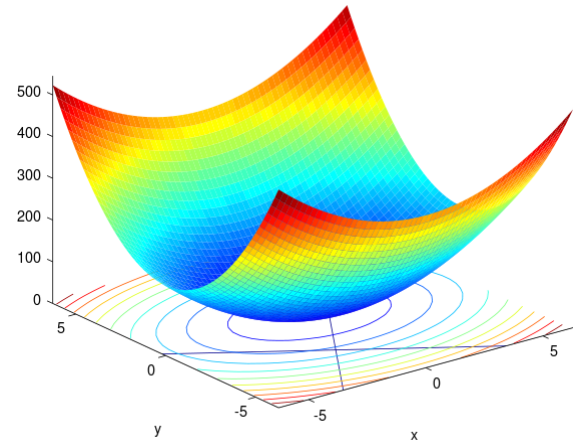

Figure 2: A paraboloid constrained along two intersecting lines.

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 a given constraint function g_i has a space of allowable directions: the space of vectors perpendicular to $\nabla g_i(\mathbf{x})$. 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.
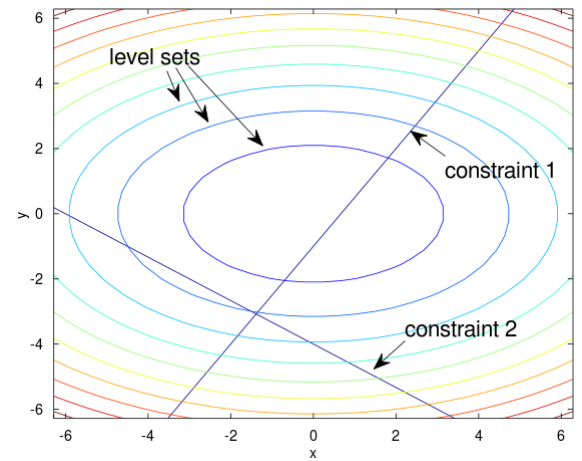

Figure 3: Contour map of Figure 2.

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

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

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

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

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

The constraint qualification assumption when there are multiple constraints is that the constraint gradients at the relevant point are linearly independent.

Modern formulation via differentiable manifolds

The problem of finding the local maxima and minima subject to constraints can be generalized to finding local maxima and minima on a differentiable manifold M.[14] In what follows, it is not necessary that M be a Euclidean space, or even a Riemannian manifold. All appearances of the gradient ∇ (which depends on a choice of Riemannian metric) can be replaced with the exterior derivative d.

Single constraint

Let M be a smooth manifold of dimension m. Suppose that we wish to find the stationary points x of a smooth function $f : M \to \mathbb{R}$ when restricted to the submanifold N defined by $g(x) = 0$, where $g : M \to \mathbb{R}$ is a smooth function for which 0 is a regular value.

Let df and dg be the exterior derivatives. Stationarity for the restriction $f|_N$ at $x \in N$ means $d(f|_N)_x = 0$. Equivalently, the kernel $\ker(df_x)$ contains $T_x N = \ker(dg_x)$. In other words, df_x and dg_x are proportional 1-forms. For this it is necessary and sufficient that the following system of $m(m-1)/2$ equations holds:

$$df_x \wedge dg_x = 0 \in \Lambda^2(T_x^* M)$$

where $\wedge$ denotes the exterior product. The stationary points x are the solutions of the above system of equations plus the constraint $g(x) = 0$. Note that the $\frac{1}{2}m(m-1)$ equations are not independent, since the left-hand side of the equation belongs to the subvariety of $\Lambda^2(T_x^* M)$ consisting of decomposable elements.

In this formulation, it is not necessary to explicitly find the Lagrange multiplier, a number λ such that $df_x = \lambda \, dg_x$.

Multiple constraints

Let M and f be as in the above section regarding the case of a single constraint. Rather than the function g described there, now consider a smooth function $G : M \to \mathbb{R}^p \, (p > 1)$, with component functions $g_i : M \to \mathbb{R}$, for which $0 \in \mathbb{R}^p$ is a regular value. Let N be the submanifold of M defined by $G(x) = 0$.

x is a stationary point of $f|_N$ if and only if $\ker(df_x)$ contains $\ker(dG_x)$. For convenience let $L_x = df_x$ and $K_x = dG_x$, where dG denotes the tangent map or Jacobian $TM \to T\mathbb{R}^p$. The subspace $\ker(K_x)$ has dimension smaller than that of $\ker(L_x)$, namely $\dim(\ker(L_x)) = n - 1$ and

$\dim(\ker(K_x)) = n - p.$ $\ker(K_x)$ belongs to $\ker(L_x)$ if and only if $L_x \in T_x^* M$ belongs to the image of $K_x^* : \mathbb{R}^{p*} \to T_x^* M$. Computationally speaking, the condition is that L_x belongs to the row space of the matrix of K_x, or equivalently the column space of the matrix of K_x^* (the transpose). If $\omega_x \in \Lambda^p(T_x^* M)$ denotes the exterior product of the columns of the matrix of K_x^*, the stationary condition for $f|_N$ at x becomes

$$L_x \wedge \omega_x = 0 \in \Lambda^{p+1}\left(T_x^* M\right)$$

Once again, in this formulation it is not necessary to explicitly find the Lagrange multipliers, the numbers $\lambda_1, \ldots, \lambda_p$ such that

$$df_x = \sum_{i=1}^{p} \lambda_i d(g_i)_x.$$

Interpretation of the Lagrange multipliers

Often the Lagrange multipliers have an interpretation as some quantity of interest. For example, by parametrising the constraint's contour line, that is, if the Lagrangian expression is

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

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

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

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 λ_k^* is the marginal cost of the constraint, and is referred to as the shadow price.[15]

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.[6][16]

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 $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, and that the minimum occurs at $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

Illustration of the constrained optimization problem 1a

For the method of Lagrange multipliers, the constraint is

$$g(x, y) = x^2 + y^2 - 1 = 0,$$

hence

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

a function that is equivalent to $f(x, y)$ when $g(x, y)$ is set to 0.

Now we can calculate the gradient:

$$\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = \left(\frac{\partial \mathcal{L}}{\partial x}, \frac{\partial \mathcal{L}}{\partial y}, \frac{\partial \mathcal{L}}{\partial \lambda}\right)$$
$$= \left(1 + 2\lambda x, 1 + 2\lambda y, x^2 + y^2 - 1\right)$$

and therefore:

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

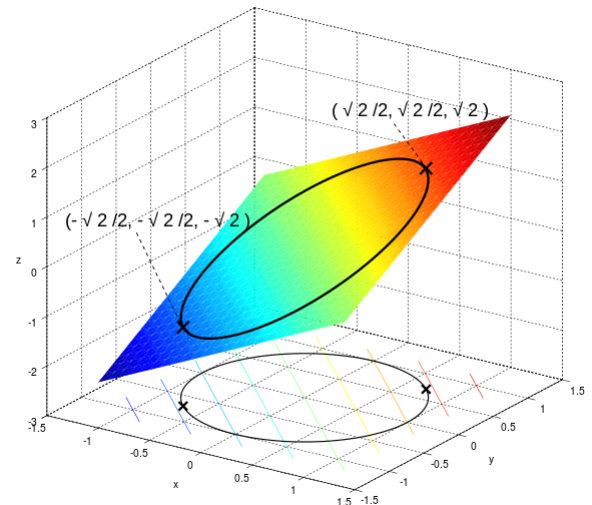Notice that the last equation is the original constraint.

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

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{\sqrt{2}}\right), \qquad \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{1}{\sqrt{2}}\right).$$

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 $g(x, y) = x^2 + y^2 - 1 = 0$. 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 $\lambda = 0$.
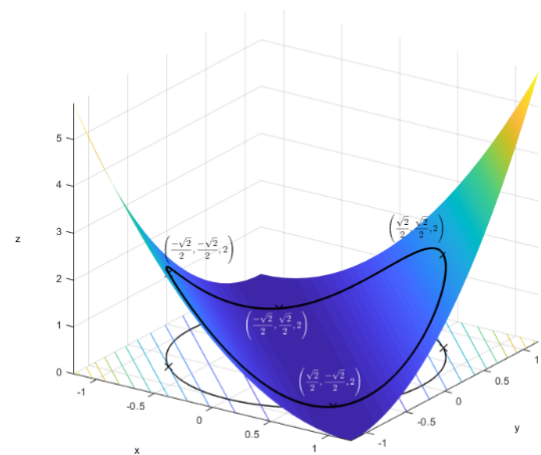

Illustration of the constrained optimization problem 1b

Example 2

This example deals with more strenuous calculations, but it is still a single constraint problem.

Suppose one wants to find the maximum values of

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

with the condition that the x- and y-coordinates lie on the circle around the origin with radius $\sqrt{3}$. That is, subject to the constraint

$$g(x, y) = x^2 + y^2 - 3 = 0.$$

As there is just a single constraint, there is a single multiplier, say λ.

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

Illustration of the constrained optimization problem

Applying the ordinary Lagrange multiplier method yields

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

from which the gradient can be calculated:

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

And therefore:

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

(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 this into (ii) yields $x^2 = 2y^2$. Substituting this into (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, one finds that

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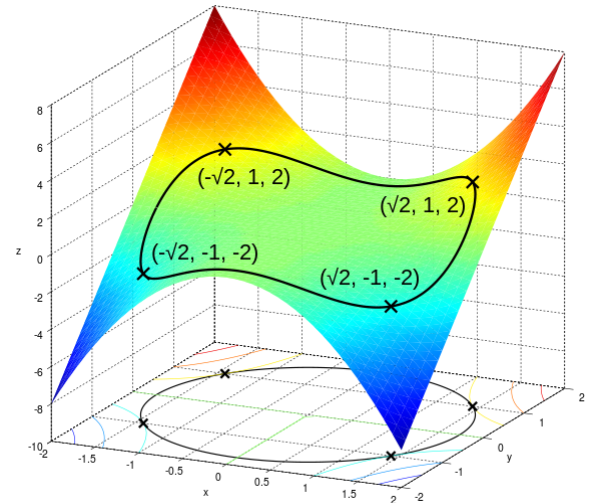
Therefore, the objective function attains the global maximum (subject to the constraints) at $(\pm\sqrt{2}, 1)$ and the global minimum at $(\pm\sqrt{2}, -1)$. The point $(0, \sqrt{3})$ is a local minimum of f and $(0, -\sqrt{3})$ is a local maximum of f, as may be determined by consideration of the Hessian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

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

Given any neighbourhood of $(\sqrt{2}, 1, -1)$, one 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 Hessian matrix of $\mathcal{L}$ evaluated at this point (or indeed at any of the critical points) which is an indefinite matrix. Each of the critical points of $\mathcal{L}$ is a saddle point of $\mathcal{L}$.[4]

Example 3: Entropy

Suppose we wish to find the discrete probability distribution on the points $\{p_1, p_2, \ldots, p_n\}$ 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:

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

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:

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

We use Lagrange multipliers to find the point of maximum entropy, $\vec{p}\,^*$, across all discrete probability distributions $\vec{p}$ on $\{x_1, x_2, \ldots, x_n\}$. 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, $k = 1, \ldots, n$, 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).[4][17] 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.

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

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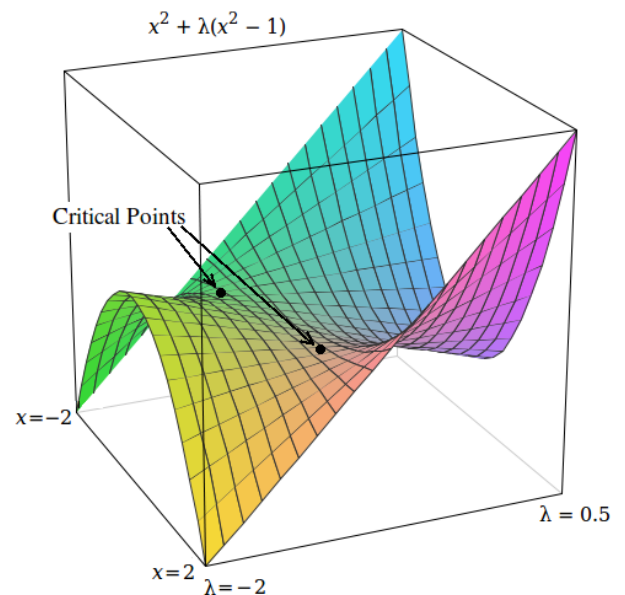
Using Lagrange multipliers, this problem can be converted into an unconstrained optimization problem:

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

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

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

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

where ε is a small value.

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:

$$h(x, \lambda) = \sqrt{(2x + 2x\lambda)^2 + (x^2 - 1)^2}$$

$$\approx \sqrt{\left(\frac{\mathcal{L}(x+\varepsilon, \lambda) - \mathcal{L}(x, \lambda)}{\varepsilon}\right)^2 + \left(\frac{\mathcal{L}(x, \lambda+\varepsilon) - \mathcal{L}(x, \lambda)}{\varepsilon}\right)^2}.$$

(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.
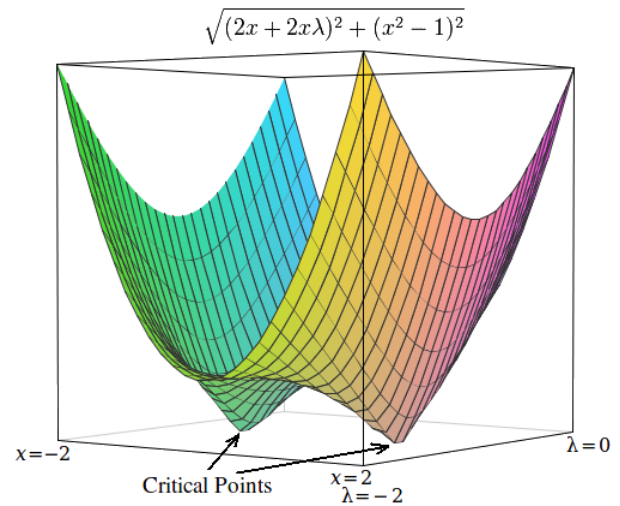
Applications

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.[18]

Power systems

Methods based on Lagrange multipliers have applications in power systems, e.g. in distributed-energy-resources (DER) placement and load shedding.[19]

See also

- Adjustment of observations
- Duality
- Gittins index
- Karush–Kuhn–Tucker conditions: generalization of the method of Lagrange multipliers
- Lagrange multipliers on Banach spaces: another generalization of the method of Lagrange multipliers
- Lagrange multiplier test in maximum likelihood estimation
- Lagrangian relaxation

References

1. Hoffmann, Laurence D.; Bradley, Gerald L. (2004). *Calculus for Business, Economics, and the Social and Life Sciences* (8th ed.). pp. 575–588. ISBN 0-07-242432-X.
2. Beavis, Brian; Dobbs, Ian M. (1990). "Static Optimization" (https://books.google.com/books?id=L7 HMACFgnXMC&pg=PA40). *Optimization and Stability Theory for Economic Analysis*. New York: Cambridge University Press. p. 40. ISBN 0-521-33605-8.
3. Protter, Murray H.; Morrey, Charles B., Jr. (1985). *Intermediate Calculus* (2nd ed.). New York: Springer. p. 267. ISBN 0-387-96058-9.
4. Walsh, G. R. (1975). "Saddle-point Property of Lagrangian Function" (https://www.google.com/boo ks/edition/Methods_of_Optimization/K0EZAQAAIAAJ?hl=en&gbpv=1&pg=PA39). *Methods of Optimization*. New York: John Wiley & Sons. pp. 39–44. ISBN 0-471-91922-5.
5. Kalman, Dan (2009). "Leveling with Lagrange: An Alternate View of Constrained Optimization". *Mathematics Magazine*. **82** (3): 186–196. doi:10.1080/0025570X.2009.11953617 (https://doi.org/1 0.1080%2F0025570X.2009.11953617). JSTOR 27765899 (https://www.jstor.org/stable/27765899). S2CID 121070192 (https://api.semanticscholar.org/Corpu sID:121070192).
6. Silberberg, Eugene; Suen, Wing (2001). *The Structure of Economics : A Mathematical Analysis* (Third ed.). Boston: Irwin McGraw-Hill. pp. 134–141. ISBN 0-07-234352-4.
7. Fuente, Angel de la (2000). *Mathematical Methods and Models for Economists* (https://archive.or g/details/mathematicalmeth00fuen). Cambridge: Cambridge University Press. p. 285 (https://archi ve.org/details/mathematicalmeth00fuen/page/n288). doi:10.1017/CBO9780511810756 (https://doi. org/10.1017%2FCBO9780511810756). ISBN 9780521585125.
8. Luenberger, David G. (1969). *Optimization by Vector Space Methods*. New York: John Wiley & Sons. pp. 188–189.
9. Bertsekas, Dimitri P. (1999). *Nonlinear Programming* (Second ed.). Cambridge, MA.: Athena Scientific. ISBN 1-886529-00-0.
10. Vapnyarskii, I.B. (2001) [1994], "Lagrange multipliers" (https://www.encyclopediaofmath.org/index. php?title=Lagrange_multipliers), *Encyclopedia of Mathematics*, EMS Press.
11. Lasdon, Leon S. (2002). *Optimization Theory for Large Systems* (Reprint of the 1970 Macmillan ed.). Mineola, New York: Dover. ISBN 0-486-41999-1. MR 1888251 (https://www.ams.o rg/mathscinet-getitem?mr=1888251).

12. 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]. Vol. 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=1295 240).

13. Lemaréchal, Claude (2001). "Lagrangian relaxation". In Jünger, Michael; Naddef, Denis (eds.). *Computational combinatorial optimization: Papers from the Spring School held in Schloß Dagstuhl, May 15–19, 2000*. Lecture Notes in Computer Science. Vol. 2241. Berlin: Springer-Verlag. pp. 112–156. doi:10.1007/3-540-45586-8_4 (https://doi.org/10.1007%2F3-540-45586-8_4). ISBN 3-540-42877-1. MR 1900016 (https://www.ams.org/mathscinet-getitem?mr=1900016).

14. Lafontaine, Jacques (2015). *An Introduction to Differential Manifolds* (https://books.google.com/bo oks?id=KNhJCgAAQBAJ&pg=PA70). Springer. p. 70. ISBN 9783319207353.

15. Dixit, Avinash K. (1990). "Shadow Prices" (https://books.google.com/books?id=dHrsHz0VocUC&p g=PA40). *Optimization in Economic Theory* (2nd ed.). New York: Oxford University Press. pp. 40–54. ISBN 0-19-877210-6.

16. Chiang, Alpha C. (1984). *Fundamental Methods of Mathematical Economics* (https://archive.org/d etails/fundamentalmetho0000chia_h4v2) (Third ed.). McGraw-Hill. p. 386 (https://archive.org/detail s/fundamentalmetho0000chia_h4v2/page/386). ISBN 0-07-010813-7.

17. Heath, Michael T. (2005). *Scientific Computing: An Introductory Survey* (https://books.google.com/ books?id=gwBrMAEACAAJ). McGraw-Hill. p. 203. ISBN 978-0-07-124489-3.

18. Pourciau, Bruce H. (1980). "Modern multiplier rules" (http://www.maa.org/programs/maa-awards/w riting-awards/modern-multiplier-rules). *American Mathematical Monthly*. **87** (6): 433–452. doi:10.2307/2320250 (https://doi.org/10.2307%2F2320250). JSTOR 2320250 (https://www.jstor.or g/stable/2320250).

19. Gautam, Mukesh; Bhusal, Narayan; Benidris, Mohammed (2020). "A Sensitivity-based Approach to Adaptive Under-Frequency Load Shedding". *2020 IEEE Texas Power and Energy Conference (TPEC)*. IEEE. pp. 1–5. doi:10.1109/TPEC48276.2020.9042569 (https://doi.org/10.1109%2FTPEC 48276.2020.9042569).

Further reading

- Beavis, Brian; Dobbs, Ian M. (1990). "Static Optimization" (https://books.google.com/books?id=L7 HMACFgnXMC&pg=PA32). *Optimization and Stability Theory for Economic Analysis*. New York: Cambridge University Press. pp. 32–72. ISBN 0-521-33605-8.
- Bertsekas, Dimitri P. (1982). *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press. ISBN 0-12-093480-9.
- Beveridge, Gordon S. G.; Schechter, Robert S. (1970). "Lagrangian Multipliers" (https://books.goo gle.com/books?id=TfhVXIWtOPQC&pg=PA244). *Optimization: Theory and Practice*. New York: McGraw-Hill. pp. 244–259. ISBN 0-07-005128-3.
- Binger, Brian R.; Hoffman, Elizabeth (1998). "Constrained Optimization". *Microeconomics with Calculus* (2nd ed.). Reading: Addison-Wesley. pp. 56–91. ISBN 0-321-01225-9.
- Carter, Michael (2001). "Equality Constraints" (https://books.google.com/books?id=KysvrGGfzq0C &pg=PA516). *Foundations of Mathematical Economics*. Cambridge: MIT Press. pp. 516–549. ISBN 0-262-53192-5.
- Hestenes, Magnus R. (1966). "Minima of functions subject to equality constraints". *Calculus of Variations and Optimal Control Theory*. New York: Wiley. pp. 29–34.
- Wylie, C. Ray; Barrett, Louis C. (1995). "The Extrema of Integrals under Constraint". *Advanced Engineering Mathematics* (Sixth ed.). New York: McGraw-Hill. pp. 1096–1103. ISBN 0-07-072206-

4.

External links

Exposition

- kipid's blog - Method of Lagrange multipliers (https://kipid.tistory.com/entry/Method-of-Lagrange-multipliers)
- 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

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)
- 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)
- MATLAB example of using Lagrange Multipliers in Optimization (http://matlab.cheme.cmu.edu/2011/12/24/using-lagrange-multipliers-in-optimization/)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Lagrange_multiplier&oldid=1070939059"

This page was last edited on 10 February 2022, at 02:44 (UTC).