Lagrange multipliers: summary

Optimisation with equality constraints: let $f:\mathbb{R}^n\to\mathbb{R}$, $\boldsymbol{g}:\mathbb{R}^n\to R^m$, $1\le m<n$. Namely, $f(\boldsymbol{x})$ is an objective function, and the notation $\boldsymbol{g}(\boldsymbol{x})=(g_1(\boldsymbol{x}),\ldots,g_m(\boldsymbol{x}))$ embraces the constraint functions, with $\boldsymbol{x}=(x_1,\ldots,x_n)$. Let $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_m)$. Consider the problem

$$\text{Min [Max]}\ \ f(\boldsymbol{x})\ \ \text{such that}\ \ \boldsymbol{g}(\boldsymbol{x})=0. \tag{1}$$

Suppose, $\boldsymbol{x}$ is a local extremiser. In addition, suppose that at $\boldsymbol{x}$, the gradients of the constraints $\nabla g_i(\boldsymbol{x})$ are linearly independent. (This is further referred to as the *non-degeneracy assumption.*) Then the constrained problem is equivalent to the unconstrained extremum problem for the *Lagrangian* (or Lagrange function)

$$L:\mathbb{R}^{n+m}\to\mathbb{R},\ \ L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})-\boldsymbol{\lambda}\cdot\boldsymbol{g}(\boldsymbol{x}), \tag{2}$$

in the following sense.

- For the latter problem, the critical points $(\boldsymbol{x},\boldsymbol{\lambda})$ must satisfy $DL(\boldsymbol{x},\boldsymbol{\lambda})=0$, i.e. solve the equations

$$\begin{aligned}\nabla f(\boldsymbol{x})-\lambda_1\nabla\boldsymbol{g}_1(\boldsymbol{x})-\ldots-\lambda_m\nabla\boldsymbol{g}_m(\boldsymbol{x}) &= 0\\ \boldsymbol{g}(\boldsymbol{x}) &= 0\end{aligned} \tag{3}$$

  Above the symbol $\nabla=D_{\boldsymbol{x}}$ means differentiation with respect to $\boldsymbol{x}$. The first group of equations in (3) is often referred to as the *Lagrange equations* (not to confuse to those in analytical mechanics), while the latter group simply repeats the *constraint equations.* So there are $n+m$ equations in (3), and there are $n+m$ unknowns. Solving these equations usually boils down to expressing $\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\lambda})$ in the first group (n equations), plugging into the second group (m equations) and getting the solutions $\boldsymbol{\lambda}$; then $\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\lambda})$.

  Indeed, if $\boldsymbol{x}$ satisfies the constraints, is a local constrained extremum, and $\boldsymbol{v}$ is a vector tangent to the feasible set at the point $\boldsymbol{x}$ – aka *true feasible direction* then the directional derivative $\nabla f(\boldsymbol{x})\cdot\boldsymbol{v}$ should be zero, or we will be able to find in the neighbourhood of $\boldsymbol{x}$ feasible points $\boldsymbol{x}'$, such that both $f(\boldsymbol{x}')>f(\boldsymbol{x})$ and $f(\boldsymbol{x}')<f(\boldsymbol{x})$. Any true feasible direction $\boldsymbol{v}$ at $\boldsymbol{x}$ has a zero dot product with any gradient $\nabla g_i(\boldsymbol{x})$, because on the feasible set $g_i(x)=0$. If the gradients $\nabla g_i(\boldsymbol{x})$ are linearly independent, they span the normal space to the feasible set at $\boldsymbol{x}$, and the Fredholm alternative to the statement

$$\nabla f(\boldsymbol{x})=\lambda_1\nabla\boldsymbol{g}_1(\boldsymbol{x})+\ldots+\lambda_m\nabla\boldsymbol{g}_m(\boldsymbol{x})$$

  is – there exists a vector $\boldsymbol{v}$, such that for all i, $\boldsymbol{v}\cdot\nabla g_i(\boldsymbol{x})=0$, while $\boldsymbol{v}\cdot\nabla f(\boldsymbol{x})\neq 0$. If the non-degeneracy assumption holds, $\boldsymbol{v}$ is a *true feasible direction* so either (3) holds or, exclusively, $\boldsymbol{x}$ is not an extremum. If it does not hold, then $\boldsymbol{v}$ coming from the Fredholm alternative is not necessarily a true feasible direction, and so the theorem may simply not work, equivalently the Lagrange equations may be inconsistent.

- After all the solutions of the Lagrange equations have been found, constrained critical points $\boldsymbol{x}$ should be characterised as local minima, maxima or saddle points, and the existence of global constrained extrema should be studied. In general this can be quite hard, however in applications it is usually straightforward. Often, for instance, there is only one $\boldsymbol{x}$ solving the Lagrange equations, and one can justify it as the global extremiser, simply because the latter should exist. To this end, the Boltsano-Weierstrass existence theorem cones handy. Besides, every problem should be thought of individually and the geometric interpretation of what's going on is vitally important, because it usually enables one to simplify the argument a lot.

  In addition, one can still make use of the second derivative test. To do so, one should *fix* $\boldsymbol{\lambda}=\boldsymbol{\lambda}$ in the Lagrangian $L(\boldsymbol{x},\boldsymbol{\lambda})$ in (2) and consider the variations of $\boldsymbol{x}$ in the vicinity of $\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\lambda})$. (If there are several values of $\boldsymbol{\lambda}$, this should be done for each $\boldsymbol{\lambda}$ and the corresponding $\boldsymbol{x}$ one after the other.) The critical point $\boldsymbol{x}=\boldsymbol{x}(\boldsymbol{\lambda})$ will be a local minimizer[maximizer] for $L(\boldsymbol{x},\boldsymbol{\lambda})$, provided that

$$D^2_{\boldsymbol{xx}}L(\boldsymbol{x},\boldsymbol{\lambda})\ >\ [<]\,0. \tag{4}$$

  The latter expressions mean, that the matrix of the second derivatives with respect to $\boldsymbol{x}$ is required to be positive [negative] definite at $\boldsymbol{x}=\boldsymbol{x}$. But the same condition is sufficient to have a local minimizer [maximizer] for the original constrained problem (1). For if $\boldsymbol{x}$ satisfies the constraint equations $\geq (\boldsymbol{x}) = 0$, then $f(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\lambda})$, and if $\boldsymbol{x}$ is the local minimizer [maximizer] for all $\boldsymbol{x}$ sufficiently close to $\boldsymbol{x}$ for the function L, it is the local minimizer [maximizer] for a narrower class of those $\boldsymbol{x}$, which satisfy the constraint equations.

- Regarding the above non-degeneracy assumption, one can easily come up with examples, failing the Lagrange method if non-degeneracy is not satisfied. E.g., in two dimensions, the constraint $g(x,y) = x^2+y^2 = 0$ makes the feasible set only one point $(0,0)$, where $\nabla g(0,0) = 0$, and in a sense any f would have a constrained extremum there. Or, take the problem $\min x$, such that $g(x,y) = x^3 - y^2 = 0$. The constraint (draw it!) implies that $x \geq 0$, and the obvious minimum for $f(x,y) = x$ is at the origin $(0,0)$, where again $\nabla g(x,y) = 0$. It is easy to see (do it!) that at $(0,0)$, the Lagrange equations, formally written, are inconsistent.

  In higher dimensions when there is more than one constraint, it is important for the Lagrange method to make sense that the gradients of the constraints $\nabla g_i(\boldsymbol{x})$ be linearly independent. Otherwise, in dimension three, with coordinates (x,y,z) take the constraints $z = 0$ and $z = y^2$. Clearly, they are satisfied by any $(x,0,0)$, i.e. on the x axis, which is the feasible set. So any objective function f that depends only on (y,z), but not on x will be constant there. I.e. any $f(y,z)$ will have a local extremum at every point in the feasible set. The Lagrange method however would require that the gradient of f be directed along the z-axis, i.e. $f_y = 0$. This is clearly not the case for any $f = f(y,z)$. Hence, in this case, the Lagrange equations will fail, for instance, for $f(x,y,z) = y$.

- Assuming that the conditions of the Lagrange method are satisfied, suppose the local extremiser $\boldsymbol{x}$ has been found, with the corresponding Lagrange multiplier $\boldsymbol{\lambda}$. Then the latter can be interpreted as the shadow price of the constraint vector. Namely, the component λ_i, equals the shadow price of the constraint $g_i(\boldsymbol{x}) = 0$ as follows.

  Suppose the constraints are written as $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{b}$ for some $\boldsymbol{b} \in \mathbb{R}^m$. Suppose, the minimizer $\boldsymbol{x}(\boldsymbol{b})$ was found for each value of $\boldsymbol{b}$, with the corresponding Lagrange multiplier $\boldsymbol{\lambda}(\boldsymbol{b})$ and objective value $f[\boldsymbol{x}(\boldsymbol{b})]$. Then, the derivative $\frac{\partial f[\boldsymbol{x}(\boldsymbol{b})]}{\partial b_i}$ is the shadow price of the ith constraint, for it tells one to the first order of δb_i, how the sought objective value will change in the leading order if b_i changes to $b_i + \delta b_i$. Indeed, in view of the fact that $\boldsymbol{x}(\boldsymbol{b})$ satisfies the constraints $\boldsymbol{g}(\boldsymbol{x}(\boldsymbol{b})) = \boldsymbol{b}$, we always have have $f(\boldsymbol{x}(\boldsymbol{b})) = L(\boldsymbol{x}(\boldsymbol{b}), \lambda(\boldsymbol{b}))$, so

$$\begin{aligned}
\frac{\partial f[\boldsymbol{x}(\boldsymbol{b})]}{\partial b_i} &= \frac{\partial}{\partial b_i} L[\boldsymbol{x}(\boldsymbol{b}), \boldsymbol{\lambda}(\boldsymbol{b})] = \frac{\partial}{\partial b_i} \left(f[\boldsymbol{x}(\boldsymbol{b})] - \boldsymbol{\lambda}(\boldsymbol{b}) \cdot (\boldsymbol{g}[\boldsymbol{x}(\boldsymbol{b})] - \boldsymbol{b}) \right) \\
&= \left(\nabla f[\boldsymbol{x}(\boldsymbol{b})] - \boldsymbol{\lambda} \cdot \nabla \boldsymbol{g}[\boldsymbol{x}(\boldsymbol{b})] \right) \frac{\partial \boldsymbol{x}(\boldsymbol{b})}{\partial b_i} - \frac{\partial \boldsymbol{\lambda}(\boldsymbol{b})}{\partial b_i} \cdot (\boldsymbol{g}[\boldsymbol{x}(\boldsymbol{b})] - \boldsymbol{b}) + \lambda_i(\boldsymbol{b}) \\
&= \lambda_i(\boldsymbol{b}).
\end{aligned}$$

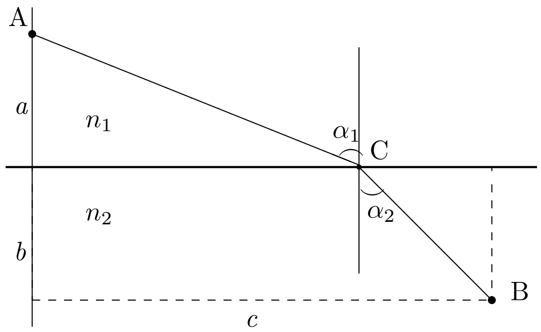
  Due to the fact that the pair $(\boldsymbol{x}(\boldsymbol{b}), \boldsymbol{\lambda}(\boldsymbol{b}))$ satisfies the Lagrange equations, the expressions in brackets in the penultimate line get annihilated! Thus, if $\boldsymbol{b} \to \delta \boldsymbol{b}$, the sought extremal value gets bigger by $\boldsymbol{\lambda}(\boldsymbol{b}) \cdot \delta \boldsymbol{b} + o(\|\delta \boldsymbol{b}\|)$, where $o(\|\delta \boldsymbol{b}\|)$ is a small error term: $\lim\limits_{\|\delta \boldsymbol{b}\| \to 0} \dfrac{o(\|\delta \boldsymbol{b}\|)}{\|\delta \boldsymbol{b}\|} = 0.$

Some examples of constrained optimisation problems

This section is optional and gives some interesting examples of how the method of Lagrange multiplies can be applied in physics and maths.

Snell's law in geometric optics

Suppose, there are two types of transparent media, separated by a thick line in the figure. Inside the first (upper) medium, the speed of light equals cn_1^{-1}, while in the second (lower) medium, it equals cn_2^{-1}, where c is the speed of light in vacuum, one of the world's constants. The number $n_{1,2}$ is called a medium's *refractive index.* For any medium, $n \geq 1$, as according to Einstein's special theory of relativity, nothing can move faster than c. For convenience, let us choose the system of units, where $c = 1$. Then in any medium, the speed of light will be smaller than 1. In the figure, $n_2 > n_1$.



The question is, given a pair of points A and B in the media 1 and 2 respectively, located as is shown in the figure, the geometrical configuration being fixed by giving the distances a, b, and c, which path will a ray of light choose to get from A to B?

The light propagates in accordance with the Fermat principle, attempting to minimize the time it takes to get from A to B. Within each [homogeneous] medium therefore, the light rays will be straight line segments; however at the surface of the media separation, the direction of a ray will change, i.e. the *light refraction* will occur. Thus, the position of the point C in the figure is unknown, and is unambiguously determined by the angles $\alpha_1, \alpha_2 < \pi/2$.

Snell's law states that one must have $n_1 \sin\alpha_1 = n_2 \sin\alpha_2$. It is easy to derive by analysing the following optimisation problem, given the quantities n_1, n_2, a, b, c:

Minimise $\frac{\frac{a}{\cos\alpha_1}}{\frac{1}{n_1}} + \frac{\frac{b}{\cos\alpha_2}}{\frac{1}{n_2}} = \frac{an_1}{\cos\alpha_1} + \frac{bn_2}{\cos\alpha_2}$, such that $a\tan\alpha_1 + b\tan\alpha_2 = c$. The objective function is just the time (the ratio of distance to speed) it takes light to get from A to B, the constraint is that the length of the horizontal projection of the line segment AB equals c.

Set up the Lagrangian:

$$L(\alpha_1, \alpha_2, \lambda) = \frac{an_1}{\cos\alpha_1} + \frac{bn_2}{\cos\alpha_2} + \lambda\left(a\tan\alpha_1 + b\tan\alpha_2 - c\right).$$

The Lagrange equations $D_{\alpha_1}L = 0,\ D_{\alpha_2}L = 0$ yield:

$$\frac{an_1\sin\alpha_1}{\cos^2\alpha_1} + \frac{\lambda a}{\cos^2\alpha_1} = 0, \quad \frac{bn_2\sin\alpha_2}{\cos^2\alpha_2} + \frac{\lambda b}{\cos^2\alpha_2} = 0,$$

and this implies $-\lambda = n_1\sin\alpha_1 = n_2\sin\alpha_2$, Snell's law.

Hadamard's inequality for determinants

Consider the following problem: let

$$A = \begin{bmatrix} x_1 & y_1 & \dots & z_1 \\ x_2 & y_2 & \dots & z_2 \\ & & \dots & \\ x_n & y_n & \dots & z_n \end{bmatrix}$$

be a square $n \times n$ matrix with n^2 unknown entries, let $\Delta = \Delta(x_1, \dots, z_1, \dots, x_n, \dots, z_n) = \det A$. Find the minimum and the maximum value of Δ, given

$$\begin{cases} x_1^2 + y_1^2 + \ldots + z_1^2 & = & h_1^2, \\ x_2^2 + y_2^2 + \ldots + z_2^2 & = & h_2^2, \\ & \dots & \\ x_n^2 + y_n^2 + \ldots + z_n^2 & = & h_n^2, \end{cases}, \quad \text{for some} \quad h_1, h_2, \dots, h_n > 0.$$

Geometrically, the ith row of A is a vector in $\mathbb{R}^n$ with the given Eucludean length h_i, $i = 1, \dots, n$. If one knows the fact that the determinant of a matrix equals the signed volume of a parallelepiped, built on its row (column) vectors, one can give the answer right away: the maximum or minimum of Δ equals $\pm\prod_{i=1}^{n} h_i$, and is achieved when all the row (column) vectors are perpendicular to one another.

Let's get the same result using the Lagrange multipliers: first of all, the above constraints ensure that the absolute extrema of the function $\Delta(x_1,\ldots,z_1,\ldots,x_n,\ldots,z_n)$ are sought on a closed bounded feasible set, and therefore exist. Set up the Lagrangian

$$L(x_1,\ldots,z_1,\ldots,x_n,\ldots,z_n,\lambda_1,\ldots,\lambda_n)=\Delta+\sum_{i=1}^{n}\lambda_i\left(x_i^2+y_i^2+\ldots+z_i^2-h_i^2\right).$$

The Lagrange equations are

$$D_{x_i}L=0,\ D_{y_i}L=0,\ldots,D_{z_i}L=0,\ i=1,\ldots,n.$$

Consider some value of i. Then for the determinant $\Delta=\Delta(x_1,y_1,\ldots,z_1,\ldots,x_n,y_n,\ldots,z_n)$, one can write the following expansion:

$$\Delta=x_iX_i+y_iY_i+\ldots+z_iZ_i,$$

where $x_i,y_i,\ldots,z_i$ are the elements of the ith row of A and $X_i,Y_i,\ldots,Z_i$ are their algebraic complements, i.e. the signed determinants of (n-1)×(n-1) matrices, obtained by deleting in A the ith row and a corresponding column. Most importantly, the quantities $X_i,Y_i,\ldots,Z_i$ do not depend on $x_i,y_i,\ldots,z_i$. So, the Lagrange equations are

$$X_i+2\lambda_ix_i=Y_i+2\lambda_iy_i=\ldots=Z_i+2\lambda_iz_i\ =\ 0,\ i=1,\ldots,n.$$

One does not need to solve these equations, note only that if $x_1,\ y_1,\ldots,z_1,\ldots,x_n,\ y_n,\ldots,z_n$ are the solutions (along with the values $\lambda_1,\ldots,\lambda_n$ for the Lagrange multipliers), comprising a constrained critical point of the function Δ, then

$$\frac{x_i}{X_i}=\frac{y_i}{Y_i}=\ldots=\frac{z_i}{Z_i}\ =\ -\frac{1}{2\lambda_i},\ i=1,\ldots,n.$$

The case $\lambda_i=0$ implies $\Delta=0$ and thus may not be considered. The uppercase symbols $X_i,Y_i,\ldots,Z_i$ have also acquired hats, as they depend on the critical values of all the rest of the variables, except $x_i,y_i,\ldots,z_i$.

Now, use some linear algebra. No matter what the variables' values are, if $i\neq j$, then

$$x_jX_i+y_jY_i+\ldots+z_jZ_i=0,$$

because this is the determinant of a matrix, obtained from A by substituting the ith row by a copy of the jth row; the determinant of a matrix, possessing a pair of identical rows is certainly zero.

Furthermore, let $B=A^TA$. Then, if A is taken at any critical point, where $\Delta\neq0$,

$$b_{ij}=x_ix_j+y_iy_j+\ldots+z_iz_j\ =\ -\frac{1}{2\lambda_i}\left(x_jX_i+y_jY_i+\ldots+z_jZ_i\right)\ =\ \begin{cases}h_i^2, & i=j,\\ 0, & i\neq j.\end{cases}$$

The case $i=j$ is just the ith constraint equation. So, at any critical point, where $\Delta\neq0$, $\det B=\prod_{i=1}^{n}h_i^2$.

Then, using the fact that $\det AA^T=(\det A)^2$, one gets immediately that at this critical point $\Delta=\pm\prod_{i=1}^{n}h_i^2$, and clearly the plus sign corresponds to the absolute minimum, and the minus sign to the absolute maximum value for the constrained problem in question. Thus, there is Hadamard's inequality:

$$\left|\det\begin{vmatrix}x_1 & y_1 & \ldots & z_1\\ x_2 & y_2 & \ldots & z_2\\ & & \ldots & \\ x_n & y_n & \ldots & z_n\end{vmatrix}\right|\ \leq\ \sqrt{\prod_{i=1}^{n}(x_i^2+y_i^2+\ldots+z_i^2)}.$$