The Method of Lagrange Multipliers — A Geometric Treatment

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

Theorem 1. *Let $g\colon \Omega \subset \mathbb{R}^{n+k} \to \mathbb{R}^k$ be C^1. Assume that $S := g^{-1}(0)$ be nonempty and that for every $p \in S$, $Dg(p)$ is of rank k. We define the tangent space T_pS of S at p to be the set of all vectors $w \in \mathbb{R}^{n+k}$ such that there exists a map c defined on some neighbourhood of $0 \in \mathbb{R}$ to S which is C^1 as a map from the interval to $\mathbb{R}^{n+k}$ with $c(0) = p$ and $c'(0) = w$. Such a map is called a C^1 curve through p in S. Thus,*

$$T_pS := \{w \in \mathbb{R}^{n+k} : \exists\ \varepsilon > 0\ \&\ c\colon (-\varepsilon, \varepsilon) \to S \subset \mathbb{R}^{n+k} \text{ such that } c(0) = p\ \&\ c'(0) = w\}.$$

We call the elements of T_pS tangent vectors to S at p. We have $T_pS =$ kernel of $Dg(p)$. In particular, T_pS is a vector space of dimension n.

Proof. It is easy to show that any tangent vector $c'(0)$ at p lies in the kernel of $Dg(p)$. For, let $c\colon (-\varepsilon, \varepsilon) \to S$ be a curve through p, i.e., a C^1 map with $c(0) = p$. The C^1-function $\sigma := g \circ c$ is then a constant so that $\sigma'(0) = 0$. By chain rule we find that $Dg(p)(c'(0)) = 0$. Thus, T_pS is contained in the kernel of $Dg(p)$.

To prove the converse, let $w \in \mathbb{R}^{n+k}$ be given such that $Dg(p)(w) = 0$. Since by hypothesis, $Dg(p)$ is of rank k, we may assume (permuting the coordinates if necessary) that $\left(\frac{\partial g_i}{\partial x_{n+j}}\right)_{1 \le i,j \le k}$ is invertible. For ease of notation, let us write for $z \in \Omega$, $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^k$. Let $p = (a, b)$ and $w = (u, v)$.

We wish to use the implicit function theorem. We repeat part of the argument of its proof to fix the notation. Let $G(x, y) := (x, g(x, y))$. Then $DG(p)$ is invertible and there exist neighbourhoods U of p in $\mathbb{R}^{n+k}$ and V of a in $\mathbb{R}^n$ and a C^1-function $h : V \to \mathbb{R}^k$ such that

(i) $h(a) = b$.

(ii) $\{z \in \Omega : G(x, y) = 0\} = \{(x, h(x)) : x \in V\}$

so that $g(x, h(x)) = 0$ for all $x \in V$. In particular, the portion $S \cap U$ of the surface S is parameterized by V.

We consider the curve $\gamma(t) := a + tu$. Since V is open for sufficiently small $|t|$, $\gamma(t) \in V$. Let $c(t) := G^{-1}(\gamma(t), 0)$. Since G is a C^1-diffeomorphism, c is C^1. Also, $c(0) = G^{-1}(\gamma(0), 0) = G^{-1}(a, 0) = (a, b) = p$.

We claim that $c'(0) = w$.

Now, by chain rule, we have $c'(0) = DG^{-1}(\gamma(0), 0) \circ \gamma'(0) = DG^{-1}(a, 0)(u, 0)$. To prove the claim, it is enough to show that $DG(p)(u, v) = (u, 0)$, as $DG^{-1}(a, 0) = DG(a, b)^{-1}$. It is easily verified that $DG(p)(u, v) = \begin{pmatrix} I_{n\times n} & 0 \\ & Dg(p) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix}$; that is

$$DG(p)(u,v) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ \frac{\partial g_1}{\partial x_1} & \dots & \dots & \frac{\partial g_1}{\partial x_n} & \frac{\partial g_1}{\partial y_1} & \dots & \frac{\partial g_1}{\partial y_k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial x_1} & \dots & \dots & \frac{\partial g_k}{\partial x_n} & \frac{\partial g_k}{\partial y_1} & \dots & \frac{\partial g_k}{\partial y_k} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ \vdots \\ u_n \\ v_1 \\ \vdots \\ \vdots \\ v_k \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix}$$

since $Dg(p)(w) = 0$. $\square$

A special case worth noting is when g takes values only in $\mathbb{R}$. Then, in that case, the condition on $Dg(p) \neq 0$ for all $p \in S$ is equivalent to requiring that $\nabla g(p) \neq 0$ for all $p \in S$. (Here $\nabla h(x) := (\frac{\partial h}{\partial x_1}, \dots, \frac{\partial h}{\partial x_n})$ for any differentiable function $h \colon U \subset \mathbb{R}^n \to \mathbb{R}$. Recall that the derivative $Dh(x) \colon \mathbb{R}^n \to \mathbb{R}$ is a linear functional given by $Dh(x)v = \langle v, \nabla h(x) \rangle$.) Note also that in this case, the vector space T_pS is the orthogonal complement of $\nabla g(p)$ in $\mathbb{R}^{n+1}$. That is why ∇g is called the normal to the "hypersurface" S.

More generally, if g is as in the theorem, we write $g = (g_1, \dots, g_k)$. We then note that $v \in T_pS$ iff $\langle v, \nabla g_i(p) \rangle = 0$ for $1 \leq i \leq k$. We say a vector $w \in \mathbb{R}^{n+k}$ is *normal* to the surface S at the point $p \in S$ iff $\langle w, v \rangle = 0$ for $v \in T_pS$. Thus, the set of vectors normal to S at p is a vector subspace of $\mathbb{R}^{n+k}$ and has $\{\nabla g_i(p) : 1 \leq i \leq k\}$.

Let M be a smooth hypersurface in $\mathbb{R}^{n+1}$ defined by $g = 0$. This means that $M := \{x \in \mathbb{R}^{n+1} : g(x) = 0\}$ and that $g : \mathbb{R}^{n+1} \to \mathbb{R}$ is a C^1-function such that $Dg(x) \neq 0$ for *any* $x \in M$. If f is a smooth function on M, we want to investigate conditions for local extremum. Let $p \in M$ be a local extremum for f. If $\gamma \colon (-\varepsilon, \varepsilon) \to M$ is a smooth curve through p, i.e., $\gamma(0) = p$, then $f \circ \gamma \colon \mathbb{R} \to \mathbb{R}$ has a local extremum at 0. Hence by calculus, we have $(f \circ \gamma)'(0) = 0$. That is,

$$\gamma'(0)(f) := \big\langle \gamma'(0), \nabla f(p) \big\rangle = 0. \tag{1}$$

Since $\gamma'(0)$ is a tangent vector to M at p, we know that

$$\big\langle \gamma'(0), \nabla g(p) \big\rangle = 0. \tag{2}$$

Now Eq. 1 and Eq. 2 imply that $\nabla f(p) = \lambda \nabla g(p)$ for some real number λ. λ is called the Lagrange multiplier. The above can very easily be generalized for submanifolds which are not necessarily hypersurfaces and which are defined by the equation $g = 0$ where g satisfies the conditions of Theorem 1. In classical language the above problem is posed as follows: Find extremum of a function f subject to the constraints $g_i(x) = 0$ for $1 \leq i \leq k$. Here $n = n + k - k$ is the dimension of the submanifold and we assume that it is defined by $g_i = 0$.

In this case the Lagrange multipliers λ_i at an extremum are given by

$$\nabla f(p) = \sum_i \lambda_i \nabla g_i(p).$$

We wish to show an interesting application of this method of Lagrange multiplier to a problem in Linear algebra or in Analytic Geometry depending on one's perspective.

Theorem 2. *Given an $n \times n$ real symmetric matrix A, there exists an orthogonal matrix U such that $U^{-1}AU = D$, where D is a diagonal matrix diag$(\lambda_1, \ldots, \lambda_n)$.*

Proof. We consider the function $g(x) := x^t Ax := \langle Ax, x\rangle$ for $x \in S := S^{n-1}$. Here $\langle , \rangle$ is the Euclidean inner product on $\mathbb{R}^n$ and

$$S^{n-1} := \{x \in \mathbb{R}^n : f(x) := \langle x, x\rangle = 1\},$$

the unit sphere in $\mathbb{R}^n$. As S^{n-1} is closed and bounded, it is compact. Hence the continuous function g attains a maximum on S. Let $v_1 \in S$ be he point where maximum is attained. We have $\nabla g(x) = 2Ax$ since $Dg(x)(h) = 2\langle Ax, h\rangle$. Also, we have $\nabla f(x) = 2Ix = 2x$. Hence by Lagrange multiplier there exists $\lambda_1 \in \mathbb{R}$ such that $\nabla g(v_1) = \lambda_1 \nabla f(v_1)$. That is, we have $Av_1 = \lambda_1 v_1$ or v_1 is an eigen value of A. Moreover, we notice that $\lambda_1 = \langle Av_1, v_1\rangle$ so that λ_1 is the maximum value of g on S.

Let $E_1 := (\mathbb{R}v_1)^\perp$, the orthogonal complement of the one dimensional subspace $\mathbb{R}v_1$ of $E_0 := \mathbb{R}^n$. We now restrict the function g to the unit sphere in E_1. That is, the variable x is constrained by $\langle x, x\rangle = 1$ and $\langle x, v_1\rangle = 0$. As above the function g attains a maximum at a point $v_2 \in S^{n-2} \subset E_1$. Then v_2 satisfies

$$Av_2 - \lambda_2 v_2 - \sigma_1 v_1 = 0 \tag{3}$$

for some real numbers λ_2 and σ_1. (Can you derive this?) We take inner product of both sides of Eq. 3 with v_2 to get:

$$\langle Av_2, v_2\rangle - \lambda_2 - \sigma_1 \langle v_1, v_2\rangle = 0.$$

Thus $\lambda_2 = \langle Av_2, v_2\rangle$ so that λ_2 is the maximum of g on the unit sphere of E_1. Also, we have $\lambda_1 \geq \lambda_2$. We now take inner product of Eq. 3 with v_1 to get

$$\langle Av_2, v_1\rangle - \lambda_2 \langle v_2, v_1\rangle - \sigma_1 \langle v_1, v_1\rangle = 0.$$

Hence we deduce that

$$\sigma_1 = \langle Av_2, v_1\rangle = \langle v_2, Av_1\rangle = \langle v_2, \lambda_1 v_1\rangle = 0.$$

Therefore from Eq. 3 we see that v_2 satisfies $Av_2 = \lambda_2 v_2$. Hence λ_2 is an eigenvalue of A.

We can thus proceed upto $n-1$ steps so that we get eigen vectors v_i of unit norm with eigen values λ_i for $1 \leq i \leq n-1$. We then consider g on S^0, which is a set of two unit vectors. Hence g has maximum say at v_n, with $\langle v_n, v_i\rangle = 0$ for $1 \leq i \leq n-1$. We set $\lambda_n := \langle Av_n, v_n\rangle$. Then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Notice that $\{v_i : 1 \leq i \leq n\}$ forms an orthonormal basis for $\mathbb{R}^n$. Hence we can write $Av_n = \sum_{i=1}^n a_i v_i$, where $a_i = \langle Av_n, v_i\rangle = \lambda_i \langle v_n, v_i\rangle = 0$ for $i < n$ and $a_n = \lambda_n$. Thus λ_n is also an eigen value of A. We have therefore diagonalised the symmetric matrix A. $\square$

Remark How is the above a result in analytic geometry? You can think of the set $\{x \in \mathbb{R}^n : g(x) = 1\}$ as a quadric surface. Thus what we have done above is to find the principal axes of this quadric surface.

Exercises

1) Find the maximum of $(x_1 x_2 \cdots x_n)^2$ subject to the constraint $\sum_1^n x_i^2 = 1$. Use this to show AM $\geq$ GM. This can also be derived by searching for the maximum of $g(x) = x_1 + \cdots + x_n$ with the constraint $f(x) := x_1 \cdots x_n = 1$.

2) Let $\Delta = \det(x_{ij})$ be the determinant of a real $n \times n$ matrix (x_{ij}). Let $v_i := (x_{i1}, \ldots, x_{in})$, the i-th row vector. Let $d_i^2 := \| v_i \|^2$. We indicate a proof of the following famous result of Hadamard:

$$|\Delta|^2 \leq d_1^2 \cdots d_n^2.$$

Let $r_i(x) := (\sum_j x_{ij}^2)^{1/2}$ be the norm of the i-th row vector. Let $d_i > 0$ be given. We wish to find the maximum of $f(x) := \det(x)$ subject to the constraints $g_i(x) := r_i(x)^2 - d_i^2 = 0$. Note that we have

$$f(x) = x_{i1} X_{i1} + \cdots + x_{in} X_{in}.$$

Here, as is customary, X_{ij} stands for the cofactor of x_{ij}. Thus if a is a point where a maximum is attained, then

$$\begin{aligned} \frac{\partial f}{\partial x_{ij}}(a) &= A_{ij} \\ \frac{\partial g_i}{\partial x_{ij}}(a) &= 2a_{ij} \end{aligned}$$

Thus the Lagrange conditions are

$$A_{ij} + 2\lambda_i a_{ij} = 0, \qquad \text{for all } i, j \tag{4}$$

Multiplying Eq. 4 by a_{ij} and summing w.r.t. j we get

$$\sum_j a_{ij} A_{ij} + 2\lambda_i \sum_j a_{ij}^2 = g(a) + 2\lambda_i d_i^2 = 0 \qquad \text{for all } i. \tag{5}$$

Multiplying Eq. 5 by a_{ij} we get

$$a_{ij} g(a) + 2a_{ij} \lambda_i d_i^2 = 0 \qquad \text{for all } i, j. \tag{6}$$

Combining Eq. 4 and Eq. 6 we get

$$a_{ij} g(a) = A_{ij} d_i^2. \tag{7}$$

We take $b := (b_{ij})$ where $b_{ij} := A_{ij}$. Then $ab = g(a)I$ so that taking determinants (i.e., applying g) on both sides yields:

$$g(a)^n = g(ab) = g(a)g(b) = g(a) \left(\prod_1^n \frac{g(a)}{d_i^2} \right) g(a). \tag{8}$$

Hence we get $(g(a))^2 = d_1^2 \cdots d_n^2$.

There is a very nice geometric interpretation of this result. First of all an observation. Since $A_{i1}a_{k1} + \cdots + A_{in}a_{kn} = 0$ for $i \neq k$ and since a_{ij} is proportional to A_{ij} we see that the rows of a are orthogonal:

$$a_{i1}a_{k1} + \cdots + a_{in}a_{kn} = 0 \text{ for } i \neq k.$$

Given n vectors $v_i := (x_{i1}, \ldots, x_{in})$ we can think of the determinant $\det(x) := \det(x_{ij})$ as an *oriented volume* of the parallelepiped $[v_1 \ldots v_n]$ spanned by the n vectors v_i. Let d_i be given positive constants. Hadamard's inequality says that under the restriction that $\| v_i \| = d_i$, the volume (:= the absolute value of the oriented volume) is a maximum when the vectors are orthogonal.

3) Show that the maximum area enclosed by a triangle of a given perimeter $2s$ is obtained by the equilateral triangle. Hint: Use Heron's formula:

$$A^2 = s(s-a)(s-b)(s-c).$$

4) Find the extrema of $f(x,y) = x^2 - y^2$ on $S := \{x^2 + y^2 = 1\}$. Draw pictures and understand the geometric meaning of your solution.

5) Show that for any $a \in \mathbb{R}^n$, we have

$$\| a \| = \max\{a.x : \| x \| = 1\}.$$

6) Let $f(x) := x_1 \cdots x_n$. Find its extrema on

$$S := \{x \in \mathbb{R}^n : \sum_k x_k = 1, \text{ and } x_k \geq 0 \text{ for all } k\}.$$

Remark 3. This may be read in conjunction with my article on "Implicit Function Theorem."